POISSON MIXTURE METHODS AND CHANGE POINT ANALYSES TO STUDY THE RELATIONSHIP BETWEEN TEMPORAL PROFILES OF SUDDEN INFANT DEATH SYNDROME AND CLIMATE

A thesis
submitted in partial fulfilment
of the requirements for the
Degree of
Doctor of Philosophy
in Statistics
by
Michelle Dalrymple

University of Canterbury
2004

For Roz

Abstract

Sudden infant death syndrome (SIDS or cot death) is the leading cause of infant death in the developed world (Byard et al., 1996), and is unique in the medical lexicon in that it represents a pathology defined by categories of exclusion. Historically, Canterbury has had one of the highest rates of SIDS in New Zealand, and New Zealand has had one of the highest SIDS rates in the western world (Nelson, 1996). SIDS is one of the most catastrophic events that can occur within a family, and is particularly traumatic due to parents invariably placing blame upon themselves for their infant's death.

Historically, statisticians and epidemiologists have played a major role in SIDS research. Their epidemiological approach has resulted in the identification of modifiable risk factors, including prone sleeping, smoking in an infant's environment, formula feeding, and bed sharing. Public health 'back-to-sleep' campaigns, have directly resulted in a sharp drop in SIDS incidence from the early 1990s.

The aim of this thesis is to model and predict the incidence of SIDS in Canterbury, New Zealand (1968—1999), in terms of complex weather patterns, characterised by a diverse array of climatic variables. This is achieved by linking the temporal sequence of SIDS counts with a comprehensive climate profile.

The association between climate and SIDS has a long history, with the first reference to a seasonal variation in SIDS (a peak in the incidence of SIDS in colder months) published nearly 150 years ago (Wakley, 1855). Many studies have related SIDS to various meteorological measures throughout the world, yet the only consistent relationship found is between SIDS and seasonality (for example Douglas et al. (1998) or Mitchell et al. (1999)). This study is the first to systematically analyse a multiplicity of climate data at different temporal levels, with an accurate extensive time series of SIDS.

Results from change point analyses showed that the Canterbury SIDS profile was constructed of three distinct temporal periods. Logistic regression on seasonality measures, confirmed that seasonality existed in the Canterbury profile. This annual variation in the incidence of SIDS was best measured by different variables for the different periods. Short term relationships between the incidence of SIDS and climate, over and above seasonality, were found for various climatic profiles including humidity, wind (speed, direction and velocity) and pressure. These relationships were identified using regression techniques based on the Poisson distribution, including Poisson regression, Poisson mixture models, Poisson regression with an autoregressive latent structure, and generalised additive models.

Three separate aspects of this study have not previously been seen in the literature, and result from novel statistical applications of mixture and change point methods to the incidence of SIDS. Firstly, this study applies mixture methods to a temporal sequence of discrete SIDS counts. Secondly, the study identifies significant points of change in the chronological profile of SIDS counts, which correspond to structural shifts in the underlying distribution. Thirdly, this study methodologically analyses a vast array of climate data at different temporal levels, with an accurate extensive time series of SIDS.

Acknowledgements

Many thanks to my family and friends for the help and support they provided me throughout the thesis process. A special mention goes to my husband Mark Dalrymple, and son Troy Butler, my mum and sisters Clare, Leanne and Josie Butler, my supervisor Irene Hudson and associate supervisor Rodney Ford, my friends Robin Turner and David Byatt, and meteorologist Alan Ryan.

Papers Published Related to this Study

Submitted Journal Articles

M.L. Dalrymple, I. L. Hudson and M.J. Faddy
Mixture Models for Time Series of Poisson counts: *with application to modelling climate and sudden infant death Biostatistics*

Refereed Journal Articles

M.L. Dalrymple, I.L. Hudson and R.P.K. Ford (2003)
Finite Mixture, Zero-inflated Poisson and Hurdle models with application to SIDS
Computational Statistics and Data Analysis, 41(3-4), 491-504.

Refereed Conference Proceedings

M.L. Dalrymple, I.L. Hudson and A.G. Barnett (2001)
Survival, block bootstrap and mixture methods for detecting change points in discrete time series data with application to SIDS
In: *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling*,
B. Klein and L. Karsholm (Eds.), 135-146.

List of Abbreviations

ACF	Autocorrelation Function
EPP	Extended Poisson Process
FM	Finite Mixture
GAM	Generalised Additive Model
GLM	Generalised Linear Model
ICD	International Classifications of Disease
LR	Likelihood Ratio
MLE	Maximum Likelihood Estimate
NB	Negative Binomial
NPMLE	Non-Parametric Maximum Likelihood Estimator
NZMS	New Zealand Meteorological Service
OR	Odds Ratio
PC	Principal Component
PCA	Principal Component Analysis
SAD	Seasonal Affective Disorder
SIDS	Sudden Infant Death Syndrome
WHO	World Health Organisation
ZIP	Zero-Inflated Poisson

Notation

The following general notation is used throughout this thesis to denote the climate variables.

Daily

Full details can be found in Chapter 1.8, page 25 onwards.

$$X_{(summary-measure)(day-span)}$$

where X represents the following climate variables:

Temperature	*Temp*	Pressure	*Pres*
Wind Direction	*WindD*	Rainfall	*Rain*
Wind Speed	*WindS*	Sunshine	*Sun*
Wind Velocity	*WindV*	Solar Radiation	*Rad*
Wind Chill	*WindC*	Dewpoint	*Dew*
Humidity	*Humid*		

Summary measures correspond to:

minimum	*min*	standard deviation	*std*	mean AHC	*meanAHC*
mean	*mean*	range	*range*	maximum AHC	*maxAHC*
maximum	*max*	total	*total*	standard deviation AHC	*stdAHC*

where *AHC* represents the absolute hourly change, which is defined as the absolute difference between any two consecutive hourly X measures.

The various day-spans are of the form:

day$i$	summary measure of X on day i
day$(i-j)$	difference in summary measure of X from i to j days ago

Also, $meanday(-1to-i)$ represents the mean of X over the past i days,
and $diff(0-i) = meanday0 - meanday(-1to-i)$ is the difference between the mean of X on the day of interest ($day0$) and the average of X over the i days prior to $day0$.

Monthly

Full details can be found in Chapter 5, page 108 onwards.

$$X^{m}_{(summary-measure)}$$

where X^m represents the monthly analogue of the climate variables denoted previously. The following summary measures are utilised:

mean(mean)	mean over the month of the daily mean of X
mean(min)	mean over the month of the daily minimum of X
mean(max)	mean over the month of the daily maximum of X
mean(std)	mean over the month of the daily standard deviation of X
min(min)	minimum over the month of the daily minimum of X
max(max)	maximum over the month of the daily maximum of X
+diff	mean over the month of all the positive values of $X_{diffday0-7}$
−diff	mean over the month of all the negative values of $X_{diffday0-7}$

Also, $WindD^m_j$ corresponds to the % of days in a month where the predominant wind direction was class j. The ten classes, j are: C calm; N north; NE north-east; E east; SE south-east; S south; SW south-west; W west; NW north-west; M no predominant wind direction (mixed).

Contents

The Ballad of Moll Magee

Come round me, little childer;
There, don't fling stones at me
Because I mutter as I go;
But pity Moll Magee.

My man was a poor fisher
With shore lines in the say;
My work was saltin' herrings
The whole of the long day.

And sometimes from the Saltin' shed
I scarce could drag my feet,
Under the blessed moonlight,
Along the pebbly street.

I'd always been but weakly,
And my baby was just born;
A neighbour minded her by day,
I minded her till morn.

I lay upon my baby;
Ye little childer dear,
I looked on my cold baby
When the morn grew frosty and clear.

A weary woman sleeps so hard!
My man grew red and pale,
And gave me money, and bade me go
To my own place, Kinsale.

He drove me out and shut the door.
And gave his curse to me;
I went away in silence,
No neighbour could I see.

The windows and the doors were shut,
One star shone faint and green,
The little straws were turnin round
Across the bare boreen.

I went away in silence:
Beyond old Martin's byre
I saw a kindly neighbour
Blowin' her mornin' fire.

She drew from me my story -
My money's all used up,
And still, with pityin', scornin' eye,
She gives me bite and sup.

She says my man will surely come
And fetch me home agin;
But always, as I'm movin' round,
Without doors or within,

Pilin' the wood or pilin' the turf,
Or goin' to the well,
I'm thinkin' of my baby
And keenin' to mysel'.

And Sometimes I am sure she knows
When, openin' wide His door,
God lights the stats, His candles,
And looks upon the poor.

So now, ye little childer,
Ye won't fling stones at me;
But gather with your shinin' looks
And pity Moll Magee.

From "The Ballad of Moll Maggee"
By W.B. Yeats (1865-1939)
From Collected Poems

Chapter 1

Sudden Infant Death Syndrome

One of the most catastrophic events that can occur within a family is a healthy infant dying suddenly and unexpectedly. Sudden infant death syndrome is particularly traumatic due to parents invariably placing blame upon themselves for their infant's death.

Sudden infant death syndrome (SIDS or cot death) is the leading cause of infant death in the developed world (Byard et al., 1996). In New Zealand, in 1995, sudden infant death syndrome accounted for 29.5% of all infant deaths (including early neonatal deaths, late neonatal deaths and post-neonatal deaths), followed by congenital anomalies (22.9%); in 1998 43% of post-neonatal deaths were caused by sudden infant death syndrome (Fetal and infant deaths, 1998).

1.1 Nature of the Study

The primary focus of this thesis is the statistical methods utilised to model a time series of (small) discrete counts. The data set utilised throughout this thesis consists of numbers of sudden infant deaths recorded in Canterbury over the years 1968–1999. The statistical modelling aims to relate the incidence of SIDS to climate patterns, while incorporating any statistically significant structural changes in the profile of SIDS over time.

The structure of the thesis consists of the historical background on SIDS, and an initial overview of the epidemiology of SIDS, thus providing the reader an opportunity to understand the phenomenon of SIDS. This first chapter also describes both the SIDS and climate datasets used throughout this study. The thesis then examines the SIDS series for any significant points of change, evidenced by both level and structural shifts in the SIDS series. Any significant changes need to be incorporated into further modelling. Subsequent chapters of the thesis focus on the statistical modelling of the relationship between SIDS and climate, with the results summarised in the final chapter.

1.1.1 Why SIDS and climate?

There has been a long association between SIDS and climate (Campbell, 1989, 1994; Campbell et al., 2001; McGlashan & Grice, 1983; Nelson & Taylor, 1988; Macey et al., 2000), yet there appears to have been no rigourus examination of the best seasonality variable to model the annual fluctuations evident in most countries with accurate SIDS records round the world. There has also been little research published, relating a complete climatic profile to the incidence of SIDS. The data set used in this thesis is unique in its accuracy and length, and also the geographical closeness of the climate data to the recorded incidence of SIDS. It was therefore considered a good choice for a statistical thesis, which, in the process, would allow for the

examination of the SIDS climate relationship in detail. The results from this study complement the current body of published research on this topic, while also adding to the statistical literature in the field of mixture models for time series data. Full details of the data used in this study appear in the following sections and chapters, and a literature review of previous research into the SIDS-climate relationship appears in Chapter 1.8

SIDS is unique in the medical lexicon in that it represents a pathology defined by categories of exclusion. The specific cause (or causes) of SIDS remains elusive to researchers with many areas currently being explored. By examing the effect of meteorological variables on the patterns of SIDS, new hypotheses and research directions may be generated. Previous research into the SIDS-climate relationship has highlighted the potential of overheating infants in the epidemiology of SIDS (Wells, 1997).

1.2 Definition

The definition for SIDS was first developed at an international conference on SIDS causes in 1969:

> *The sudden and unexpected death of an infant or young child, which is unexpected by history, and in which a thorough post-mortem examination fails to demonstrate an adequate cause of death* (Beckwith, 1970).

This definition was later amended in 1989 by a group assembled by the United States National Institute of Health to:

> *The sudden death of an infant under 1 year of age which remains unexplained after a complete postmortem examination, including an investigation of the death scene and a review of the case history, cases failing to meet the standards of this definition, including those without postmortem examinations, should not be diagnosed as SIDS* (Zylke, 1989).

In 1977, the World Health Organisation created a category, in the International Classifications of Disease, specifically for cot death or sudden infant death syndrome: ICD code 798.0 (World Health Organisation, 1977). The coding guidelines applied to this classification are conservative, in that if another condition is reported together with the SIDS classification, the underlying cause of death is officially recorded as the second condition. Nelson (1996) gives the following example of this scenario: "The pathologist might decide on a diagnosis of SIDS, but also records a coexistent condition (for example pneumonia, ICD code 486.0). Thus, even though the pathologist considered the death to be due to SIDS, the official statistics will not record it as such." This rule may well have a significant effect on the official statistics of SIDS incidence, with what could be considered a systematic under-reporting.

1.3 Historical Review

Traditionally, the cause of sudden infant death syndrome was assumed to be overlaying (suffocation), either through smothering by the mother, or by bedclothes. Until the first quarter of last century, bed-sharing between parents and young infants was very common (with the exception of the upper classes of society) (Knight, 1983). Thus, when an infant died suddenly and unexpectedly, and was found in the family bed, the naturally drawn conclusion was that the infant had suffocated by being buried or overlayed.

The earliest known reference to a possible sudden and unexpected infant death appears in the Bible: "and this woman's child died in the night because she overlaid it" (Kings I, chapter 13, verse 19). At a

similar time in Egypt, Diodurus Siculus reported that the punishment given to mothers who were believed to have overlain their infants was to hold their child continually for three days and nights "so that the mother would experience her full deserts of remorse and horror" (Russell-Jones, 1985).

When the prevailing theory explaining sudden infant deaths was overlaying, infanticide was also common and few endeavours were made to distinguish between the two disparate causes of death (Golding et al., 1985). Apparently, the early Roman Catholic church was not actively against infanticide, and both overlaying and infanticide were only considered sinful, though not criminal, by the church around 700AD (Guntheroth, 1995).

In the nineteenth century, research into general infant mortality sparked increasing interest into the phenomenon of sudden infant deaths. The following passage by Charles Dickens may well reflect his thoughts on infant mortality:

> *Of all things in life, there should be nothing so preventable, as there is nothing on the face of it so unnatural, as the death of a little child, yet it is of all things in life the commonest, the only thing we really make, as a community, the least effort to prevent.*

The overlaying theory was still considered the predominant cause of most sudden and unexpected infant deaths in the late nineteenth century, with the following text on infant mortality appearing in the English's Women's Journal:

> *Nor must we forget a lamentable but frequent cause of death, that in which the infant is "overlaid" in its slumbers by a careless, perhaps drunken nurse or mother. The deaths of no less than twelve infants from suffocation in bed were lately registered as having occurred within a single week.* (Leared, 1862; cited in Russell-Jones, 1985)

Nearly a decade later in 1871, a paper titled '*The Waste of Infant Life*' was presented to a social science association by Curgenven. Curgenven reported statistics showing that, of all infants on whom inquests were held, more than half had died from being overlayed:

> *Suffocation at the mother's breast is the most common form: this frequently occurs on Saturday night; the mother, it may be, has been drinking, or goes to bed late; takes her child to her breast; falls asleep, and awakes in the morning to find her infant dead. In most cases the child's head slips off the mother's arm during sleep, its face becomes buried beneath the breast, the bed-clothes at the same time covering its head; it dies without a struggle suffocated by the carbonic acid exhaled from its own lungs. In winter these cases of suffocation occur most frequently in consequence of the mother unconsciously drawing the bed-clothes over her own shoulders* (Curgenven, 1871; cited in Russell-Jones, 1985)

Interestingly, this passage suggests a winter predominance of sudden and unexpected infant deaths. The first reference to a seasonal variation in SIDS occurred some 15 years earlier, when Wakley (1855) noted a peak incidence in the colder months.

In 1889, the 'status-thymico-lymphaticus' theory was presented by Paultouf as an alternative to overlaying as the cause of sudden infant death. This theory proposed that an enlarged thymus gland was responsible for death. The following account was described in an 1898 paper:

> ... *a case in which a servant girl was entrusted with the care of an infant who was sleeping in its cradle near her. In the morning the previously healthy baby was found dead in his cradle. The girl was imprisoned and authorities ordered an autopsy which was performed by Liman and Gravitz. Liman was unable to disprove Gravitz' contention that death was attributable to a colossally enlarged thymus pressing on trachea, bronchi and vessels. The magistrate released the woman* (Bertholds 1898, cited in Russell-Jones, 1985).

Paultouf's theory was later disproven when it was shown that the thymus gland in sudden death infants was in fact a normal size, but appeared large when compared to the involuted gland that occurs after a long illness.

As the thymus gland theory waned in popularity, blame for sudden infant deaths returned to those caring for the child. In 1892 Templeman presented the first systematic report of the pathology and epidemiology of sudden infant death. He reported that 399 infants were found dead while in their parents bed, in the ten year period 1882 to 1891, in the small Scottish town of Dundee. This corresponds to an 'overlaying' death rate of approximately 4.1 infant deaths per 1000 live births (Dundee birth figures from the General Register Office for Scotland) Templeman described several of the modern characteristics of SIDS: the cyclicity in the temporal distribution leading to higher rates in winter, and the typical age distribution with a peak death rate occurring in infants at three months of age. Templeman believed that the deaths were caused by overlaying or suffocation from either the infant's bedding or a parent's body. He theorised in turn, that the overlaying was caused by a lack of knowledge of the dangers of bed sharing, and also noted a disproportionately large number of sudden infant deaths happening on Saturday nights, due to frequent drunkness on Saturday evenings. Templeman made the following recommendations in his paper:

> ... *although savouring somewhat of grandmotherly legislation, we ought to adopt a modification of the German law which prohibits a child under the age of two years from being allowed to occupy the same bed as its parents or nurse ... But even short of this measure, the punishment of parents who can be proved to have gone to bed in a state of intoxication would have a beneficial effect"* (Templeman, 1892)

Templeman's theory was strengthened by Saleby in 1917 (cited in Russell-Jones, 1985) when he reported a nearly 50% reduction in infant deaths from overlaying (from 1348 deaths in 1912 to 704 in 1917). During this period a change in licensing laws, restricted "drunkenness in women", which Saleby stated was the sole reason for the reduction in deaths.

At the turn of last century, the concept of sudden infant death was becoming more widely known. Yeats' 'Ballard of Moll Magee' (1906) (page xii) poignantly illustrates society's attitudes to a mother who was believed to have overlain her infant. As the standard of living increased early last century, many infants were no longer sleeping in the family bed, so the overlaying theory became less plausible. Suffocation by the infant's bed clothes replaced the overlaying theory as the probable cause of sudden infant deaths. This was later disproven when it was shown that most infants slept in a supine position, with their faces clear of their bedclothes (Beckwith, 1973).

After the second world war, most pathologists thought that sudden infant deaths were a result of natural causes, although the exact nature of the disease remained unknown. Rather than admit that they could not ascertain the cause of death, it was common for pathologists to state that deaths were due to respiratory infections (Knight, 1983).

The first attempt to detail a natural process for the cause of sudden infant deaths came in the late 1950s. Parish & Barett (1960) proposed that the deaths were caused by an allergic reaction to a foreign protein found in cows' milk. Later studies have failed to show that this reaction occurred more frequently in SIDS infants as compared with normal infants (Gold & Godek, 1961; Coe & Peterson, 1963; Ford et al., 1996).

In 1965, an 'Enquiry into Sudden Death in Infancy' was published (Reports on Public Health, 1965), and proposed three general theories as the cause of SIDS: infection, hypersensitivity and suffocation. This publication highlighted the immense knowledge gap in the epidemiology of SIDS and detailed research priorities.

From that time on, there has been an abundance of theories attempting to explain the enigma of sudden and unexplained deaths in infants. For example apnea (Froggatt et al., 1971; Steinschneider, 1972; Toubas et al., 1986), previous hypoxic episodes (Naeye, 1973), bacterial or viral infections (Trube-Becker, 1978; Newbould et al., 1989; Knox & Lancashire, 1991; Williams et al., 1984; Zink et al., 1987), hyperimmune reactions (Raven, 1977; Raven et al., 1978), sickle cell anaemia (Vix et al., 1987; Gozal et al., 1994), toxic gases from cot mattresses (Richardson, 1990) and hyperthermia (Stanton, 1984; Kleemann et al., 1996).

Only in the last 10 to 15 years have the first major inroads been made into reducing SIDS numbers. In the early 1990s many countries, including New Zealand (Mitchell et al., 1992a), the United Kingdom (Wigfield et al., 1994) and the United States (Pollack & Frohna, 2002), implemented public health campaigns aimed primarily at reducing the rate of prone sleeping in infants. These 'back-to-sleep' campaigns have resulted in dramatic reductions in SIDS numbers, approximately halving SIDS in most countries, and reducing numbers by up to 70% in some countries (for example Sweden, (Alm et al., 2001)). Even after such progress, SIDS still accounts for the most infant deaths in industrialised countries (Byard et al., 1996). Researchers have been able to ascertain modifiable risk factors, which in turn have been widely promoted to good effect, yet the exact reason why seemingly healthy infants die remains a mystery.

1.4 Worldwide Incidence

Sudden infant death syndrome is essentially a syndrome of the developed world. Before the industrial revolution SIDS statistics were obscured by very high general infant mortality rates. This presently remains the case in non-industrialised countries. For example, for the 1994 to 1997 period the overall rate of infant mortality in Japan was 4.2 deaths per 1000 live births, and in New Zealand 6.7 deaths per 1000 live births, as compared to 57 deaths per 1000 live births for Brazil and 77 deaths per 1000 live births for Bangladesh (World Health Organisation, 2002b). Any SIDS rate is swamped by the mortality rate in the third world due to disease and poverty. In many countries in the developing world, infectious diseases such as pneumonia, diarrhoea and malaria account for up to 70% of deaths in children under five years old (World Health Organisation, 2002a).

Figure 1.1 presents SIDS rates for various countries. The incidence rates are the number of SIDS deaths per 1000 live births in the late 1990s, as reported by the World Health Organisation (WHO), or the SIDS Global Strategy Task Force (Fitzgerald, 2000). Appendix A tabulates full details of SIDS rates by country, and the source of data (where numbers were available from the Task Force, these were used in the figure). In the industrialised world, the differing SIDS rates, whilst having some relevance, must also be tempered by the non-standardised reporting practices used. The diagnosis of SIDS formed by exclusion presents difficulties in data accuracy. Therefore any international comparative analysis, whilst still of interest, must be viewed with caution. Also note, by using the WHO SIDS rates, numbers may well be higher than presented here, as discussed in Section 1.2.

In different countries SIDS rates currently vary from less than 0.1 per 1000 live births in Hong Kong and Hungary (1998 figures) to around 1.0 per 1000 live births in New Zealand (see Table A.1). New Zealand's

very high SIDS rate was evident in the 1980s before the 'back-to-sleep' campaigns, with 4.2 SIDS per 1000 live births in 1982 (Borman et al., 1988). High rates were also reported at this time in Australia (2.0/1000 LB, Douglas et al. (1996)), England and Wales (2.0/1000 LB, Campbell (1994)), Ireland (2.4/1000 LB, Matthews & O'Brien (1985)) and Scotland (2.7/1000 LB, Arneil et al. (1985)). Hong Kong (0.3/1000 LB, Lee et al. (1989)), Taiwan (0.5/1000 LB, Knöbel et al. (1994)), Finland (0.5/1000 LB, Rintahaka & Hirvonen (1986)) and Sweden (0.9/1000 LB, Milerad et al. (1993)) recorded very low rates around this period.

1.5 The Epidemiology of SIDS

A multitude of epidemiological studies have identified various risk factors of SIDS, resulting in a dramatic reduction in SIDS incidence, yet the exact cause of sudden infant death syndrome remains unknown despite the fact that more than 400 hypotheses have been proposed. The original search for a single cause has evolved into an understanding that SIDS has a multifactorial aetiology, with a complex underlying interaction of risk factors. Toomey & Bernstein (2001) suggest that it is this interaction of risk factors, specifically "[the] developmental stage, congenital or acquired risk, along with a triggering event" that eventually ends in SIDS. Nelson (1996) noted that "for a condition that is defined on the basis of exclusion, the epidemiological features of SIDS are surprisingly consistent".

The following section describes some of the major epidemiological characteristics and risk factors of SIDS. It does not attempt to detail any of the many hypotheses and trigger mechanisms put forward to explain them. For a detailed review, see Sullivan & Barlow (2001), or Guntheroth (1995).

1.5.1 Characteristics of SIDS Epidemiology

Age-at-Death Distribution

In 1980, Peterson described the age-at-death of SIDS infants as "the single most consistent, provocative and unique characteristic of SIDS yet identified". Figure 1.2 illustrates the classical age-at-death distribution of SIDS infants, for deaths occurring in King County, Washington, 1969—1977. The frequency distribution shows that very few SIDS deaths occur within the first month of life. This is the time of maximum frequency of infant deaths from most other causes (Goldberg et al., 1986). The distribution curve rises steeply to a peak age of vulnerability at three months, and then rapidly decays to six months, after which age, only a small number of deaths occur. This pattern in the age-at-death frequency distribution of sudden and unexpected infant deaths has been noted for over a century, for example see Templeman's report (1892).

Adams et al. (1998) examined the age-at-death distribution among Californian infants after the implementation of 'back-to-sleep' campaigns. They reported that for black infants, the proportion of SIDS deaths between the second and fourth months of life (the peak age period of deaths) remained unchanged post 'back-to-sleep' campaigns, while, for non-black infants, the proportion of deaths during the second to fourth months decreased significantly. In both groupings, the peak age-at-death remained the third month of life. In a study examining SIDS epidemiology in England between 1993 and 1996, Leach et al. (1999) found the characteristic age distribution remained the same after the 'back-to-sleep' campaigns as it was before.

Figure 1.1: Sudden infant death syndrome international rates (1992—2000).

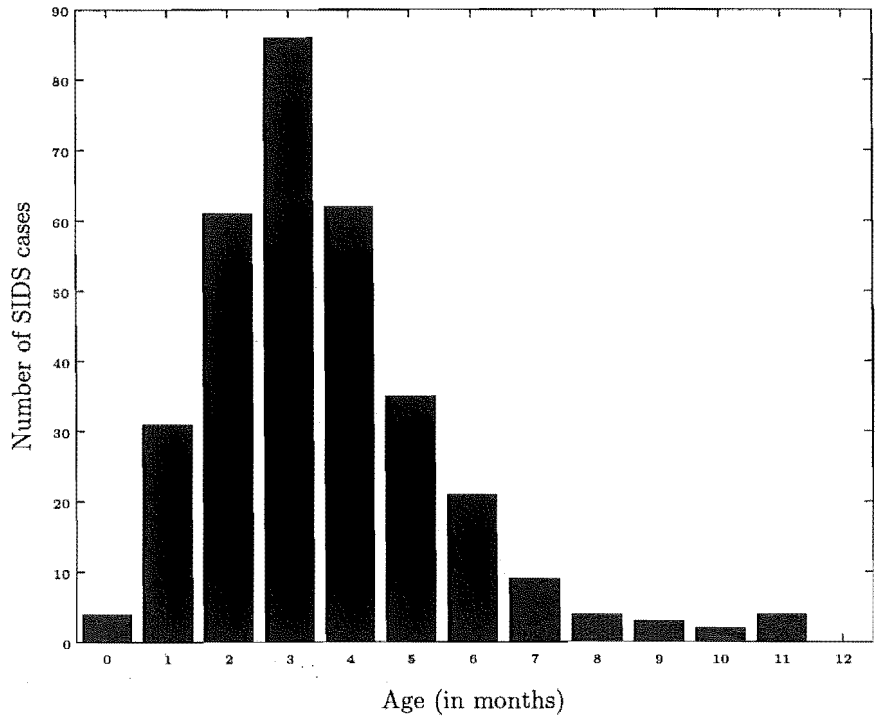


Figure 1.2: Sudden infant death syndrome (SIDS) age-at-death distribtuion, King County, Washington, 1969—1977 (Source: Peterson 1980).

Seasonal Distribution

Prior to the 'back-to-sleep' campaigns, the seasonal pattern of SIDS deaths was reported in many studies (for example Douglas et al. (1996) or Guntheroth (1995)). Figure 1.3 illustrates the annual sinusoidal variation in SIDS numbers for Canterbury, New Zealand, for the period 1973 to 1989. Peak SIDS incidence rates occurring in winter, with a corresponding low in the warmer months (mean SIDS numbers for winter over this period are 11.0 compared with a summer mean of 3.9). This characteristic of SIDS epidemiology has also been known for well over a century, with Curgenven reporting an increased winter incidence in sudden and unexpected deaths in his 1871 paper (cited in Russell-Jones (1985)).

Some studies have suggested that, with the reduction in SIDS rates post 'back-to-sleep' campaigns, the strong seasonal effect is no longer evident. Leach et al. (1999) found no winter SIDS predominance when examining SIDS in England from 1993 to 1996. In contrast, both Douglas et al. (1998) and Julious (1997) have shown a persistence of seasonality in SIDS rates in Britain post campaigns, although the seasonal amplitude reduced by approximately a half. In New Zealand, Mitchell et al. (1994) found that the reduction in SIDS incidence post 'back-to-sleep' campaigns occurred predominantly in the winter months, with little change in the SIDS rates in summer, though the winter peak still existed.

Geographical Distribution

Prior to the commencement of 'back-to-sleep' campaigns, SIDS incidence increased with increasing latitude (a north to south gradient) in both New Zealand and Australia. In the early 1980s, New Zealand's SIDS rate ranged from 2.5 per 1000 live births in Auckland (latitude 36°S), increasing to 7.3 per 1000 live births in Dunedin and Invercargill (latitude 45°S and 46°S respectively) (Nelson, 1996). Figure 1.4 illustrates

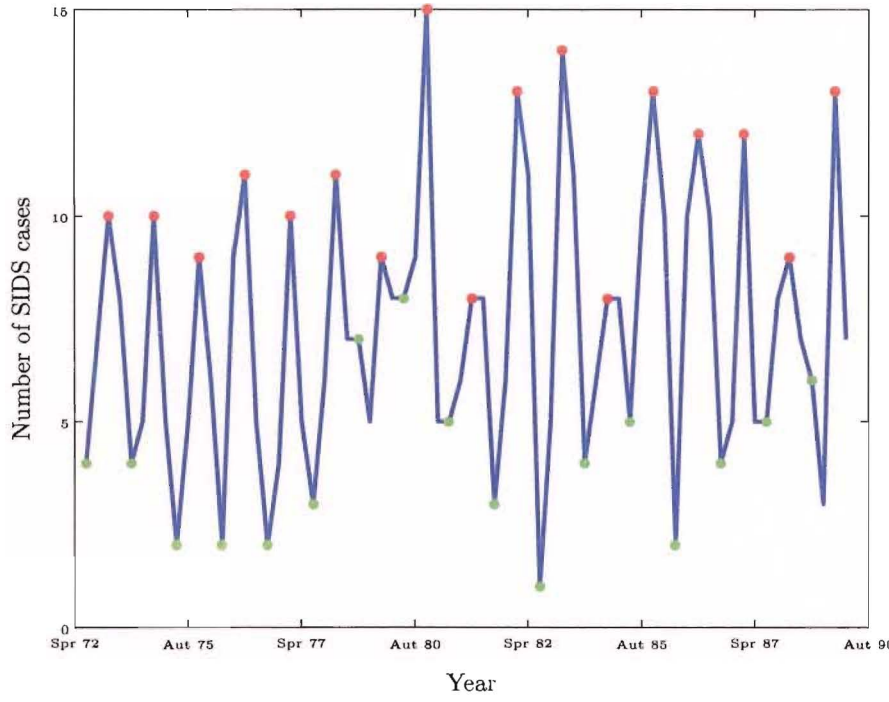


Figure 1.3: Sudden infant death syndrome numbers, by season, for deaths occurring in Canterbury, New Zealand, 1973—1989 (•= summer; •= winter).

this gradation of SIDS rates throughout New Zealand. Similarly in Australia, SIDS rates for Queensland (latitude 29°S to 12°S) and Tasmania (latitude 41°S to 43°S) were 1.6 SIDS per 1000 live births and 3.5 SIDS per 1000 live births respectively (Social Report, 1991).

In contrast, studies have not shown significant differences in the United States SIDS incidence rates with respect to latitude, although Spiers (1990) reported a longitudinal risk gradient from east to west in the United States.

The north to south gradient is no longer evident in overall New Zealand SIDS rates post 'back-to-sleep' campaigns. As shown in Figure 1.5, the overall SIDS risk for 1997, by region, appears reasonably uniform across Health districts. Yet, the risk gradient for Maori infants currently runs in the opposite direction, directly reflecting the population distribution of the Maori risk group (Fetal and Infant Deaths, 1997).

Ethnic Variations

There are ethnic and racial differences in the recorded SIDS rates. Generally, rates are highest in black American, native American, Maori and Aboriginal populations, mid-way in Caucasian populations, low in Hispanic populations and lowest in Asian populations (Sullivan & Barlow, 2001).

Some of the differences in SIDS rates between ethnic groups can be attributed to socio-economic factors. However, Borman et al. (1988) reported differences in SIDS rates for Maoris (6.5 per 1000 live births), Caucasians (3.9 per 1000 live births) and Pacific Islanders (1.9 per 1000 live births). These differences could not be explained by socio-economic factors as both Maoris and Pacific Islanders are similarly socially

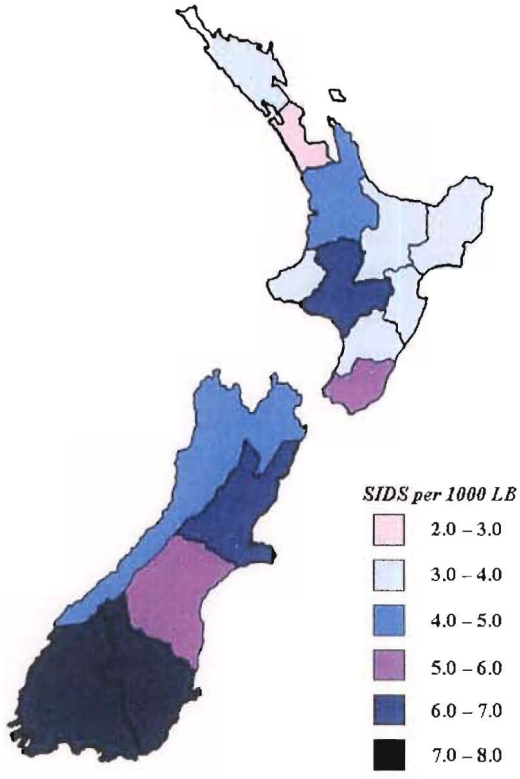
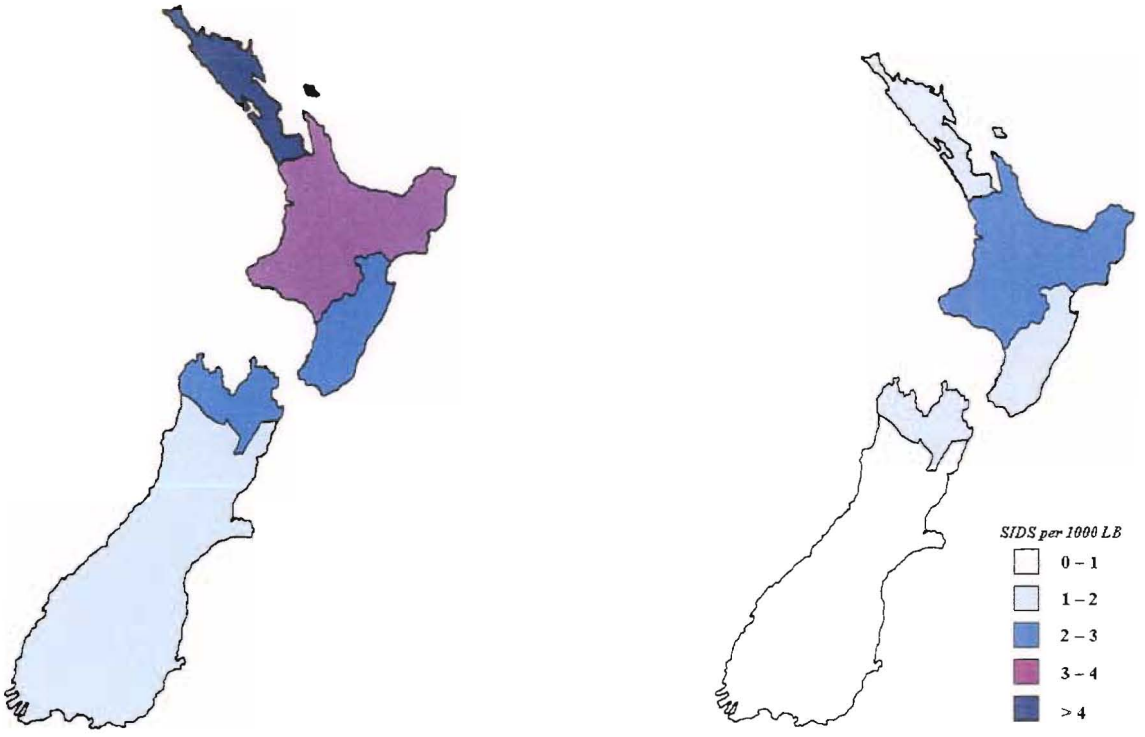


Figure 1.4: Rates of sudden infant death syndrome in New Zealand (1980—1984).



(a) Maori

(b) Total

Figure 1.5: Sudden infant death syndrome, 1997, Maori and Total rates (Regional Health Authority Regions).

	SIDS rates/1000 LB	SIDS risk	Socioeconomic status
Maori	7.9	high	low
Caucasian	1.4	medium	high
Pacific Islander	0.9	low	low

Table 1.1: SIDS risk and socioeconomic status by ethnic grouping in New Zealand (1993).

disadvantaged, yet the SIDS rate for Pacific Islanders is half that of Caucasians. This situation is detailed in Table 1.1. A similar paradoxical situation exists in the United States where the high SIDS rate in native and black American populations is not reflected in the Hispanic community, an ethnic group of similar socio-economic status. Numbers from a study in Cook County from 1975 to 1980 showed the Hispanic SIDS rate was 1.2 per 1000 neonatal survivors, more than 75% lower than that of blacks in the United States (5.1 SIDS per 1000 neonatal survivors) and lower still than the Caucasian rate of 1.3 SIDS per 1000 neonatal survivors (Black et al., 1986).

Grether & Schulman's study (1989) showed that, in California, Hispanic SIDS rates varied with the mothers' country of origin: the SIDS rate for mothers born in the United States was 1.5 per 1000 live births as compared to 0.8 per 1000 live births for mothers born in Mexico. These results suggest that an important determinant of SIDS may be ethnic variations in infant care practices (Dwyer & Ponsonby, 1992).

Although SIDS numbers have dropped dramatically after the advent of 'back-to-sleep' campaigns, they have not dropped uniformly across different ethnic communities. For example, in New Zealand from 1989 to 1996, the Maori SIDS rate dropped by 57% and the Pacific Islander rate by 52%, while the SIDS rate for other ethnic groupings dropped by 82% (Fetal and Infant Deaths, 1997).

1.5.2 SIDS Risk Factors

Table 1.2 presents epidemiological, behavioural and environmental risk factors which have been identified by a number of studies as associated with SIDS risk. It is important to be aware that not all studies confirm a relationship between all variables presented, though a majority support the factors described in the following section. It is also important to note that there is an interdependence between some of the presented risk factors, for example a premature infant is likely to have had a low birth weight, yet both prematurity and low birth are well known risk factors. These two factors are in turn influenced by maternal smoking, another risk factor of SIDS (Guntheroth, 1995).

Maternal Risk Factors

Many studies have found that SIDS incidence increases in young mothers, particularly mothers under 20 years of age (Dwyer & Ponsonby, 1995). This association was first reported by Steele & Langworth in 1966. SIDS incidence varies inversely with maternal age, in contrast to general postneonatal mortality, which has a parabolic relationship with maternal age. Several studies have found a higher SIDS risk for unmarried mothers, with Borman et al. (1988) reporting the SIDS risk for unmarried mothers was 70% higher than for married mothers.

A significant relationship between lower socio-economic status and SIDS has been documented in many studies (Guntheroth, 1995). The degree to which socio-economic status is a proxy for other risk variables, or vice versa, is a constant problem. In a study of SIDS in New Zealand, Borman et al. (1988) reported a significant inverse relationship between SIDS incidence and socio-economic status, with rates ranging from 3.0 SIDS per 1000 live births in the highest income group, to nearly double this at 5.7 per 1000 live births

Increased SIDS risk	**Decreased** SIDS risk
• Maternal factors	• Postnatal factors
Young age	Sleep position
Unmarried	Supine
Lower socio-economic status	Side
• Antenatal factors	Bed sharing
Poor attendance at antenatal clinic	–With non-smoking adult
Prematurity	Room sharing (with adults)
Maternal smoking	Breastfeeding
Maternal caffeine consumption	Pacifier use
• Birth characteristics	Immunisation
Male infant	
Low birth weight	
• Postnatal factors	
Sleep position	
–Prone	
Soft bedding/sleep surface	
Bed sharing	
With smoking adult	
Overheating	

Table 1.2: Some selected risk factors associated with SIDS.

for the lowest income group. In a recent case-control study of SIDS incidence in England post 'back-to-sleep' campaigns, Leach et al. (1999) showed an altered socio-economic context of SIDS as compared to the socio-economic distribution of SIDS prior to the 'back-to-sleep' campaigns. Their study showed a discernible increase of lower income families in the SIDS group (63% SIDS compared to 27% controls), with half of the SIDS families in the study being unemployed. The inverse relationship between maternal age and SIDS can, in part, be explained by socio-economic status, but not entirely so. For example, a study by the National Institute of Child Health and Development (NICHD) examining SIDS incidence, reported an unadjusted odds ratio (OR) of 3.5 for a teenage mother. This dropped to 2.4 after adjusting for socio-economic status (Harper & Hoffman, 1988).

Antenatal Risk Factors

Certain pregnancy characteristics influence an infant's risk of sudden death. The risk of SIDS increases when mothers have a poor attendance record at antenatal clinics (Golding et al., 1985; Peterson, 1980). SIDS risk also increases with a shorter gestational period, a factor that is intricately linked with low birth weight, also a significant risk factor (Guntheroth, 1995). Borman et al. (1988) reported a SIDS rate of 11.15 per 1000 live births for infants with a gestational age of 28-33 weeks. This was significantly different from 3.82 SIDS per 1000 live births, the SIDS incidence rate for infants with a gestational age of at least 38 weeks. The Oxford Record Linkage Survey found an increased risk of SIDS in premature (< 35 weeks) infants with an OR of 3.9 (Golding et al., 1985).

Maternal smoking during pregnancy has long been associated with increased risk of SIDS. Again, the 1966 study by Steele & Langworth was the first to present this association. Two recent studies by Pollack (2001) and Wisborg et al. (2000) have presented results showing that the SIDS risk is more than double

for infants whose mothers smoked antenatally as compared to infants of non-smoking mothers. Although maternal smoking during pregnancy is well known to be associated with low birth weights of infants, after adjustment for birth weight, there still remains a residual positive association between maternal smoking and SIDS. For example, a study by Haglund & Cnattingius (1990) found no significant reductions in the risk ratios associated with moderate and heavy smoking after controlling for low birth weight (classified as < 2500g). Results from a cohort study by Schellscheidt et al. (1997) reported no increase in SIDS risk for a premature infant when the mother was a non-smoker, whereas the risk associated with a birth of less than 37 weeks gestation was significantly increased if the mother smoked during pregnancy (relative risk of 19.6 when compared to full term infants of non-smoking mothers).

There has been little research published on the effects of passive smoking on SIDS, as it is very difficult to separate the effects of maternal smoking during pregnancy and smoking in the infants environment after birth. In a meta-analysis by Anderson & Cook (1997), 39 studies on the effects of parental smoking on SIDS were examined. The contribution of postnatal maternal smoking was estimated, after controlling for maternal prenatal smoking, with an adjusted OR of 1.94. Anderson and Cook's study highlighted the inconclusive results found for the effect on SIDS incidence of other smokers in the household, with one third of studies reporting non significant relationships. From the research presented to date, it has not been possible to tell whether the increased SIDS risk for infants born to smokers is a result of smoking during pregnancy, passive smoking postnatally, or a combination of both.

There have been few detailed studies of the effect of caffeine intake during pregnancy on the risk of SIDS. Hoffmann et al. (1988) reported that there was no effect from caffeine on SIDS risk. Similarly, Alm et al. (1999) found that, after adjusting for possible confounders, caffeine intake during or after pregnancy was not significantly associated with the incidence of SIDS. In contrast, a recent study by Ford et al. (1998) reported a significantly increased risk of SIDS for infants whose mothers were classified as heavy caffeine consumers ($\geqslant$ 400mg/day, or $\geqslant 4$ cups of coffee/day) throughout pregnancy (OR of 1.65) after adjusting for potential confounding variables.

Risk Factors at Birth

The majority of epidemiological reports indicate that SIDS incidence rates for male infants are about 1.5 times higher than for female infants. Borman et al. (1988) reported a male SIDS rate of 4.88 per 1000 live births, about 1.4 times higher than females (3.42 SIDS per 1000 live births) in his study of SIDS in New Zealand. This male predominance of SIDS is not found in all ethnicities. Borman's study noted that the SIDS rates for Maori male infants was no different to the Maori female rates, even though there was a very high overall Maori SIDS rate (Borman et al., 1988). The association of increased risk for male infants is not unique to SIDS, but is found for all other causes of infant mortality (Peterson, 1989).

Low birth weight is a strong predictor of SIDS. As mentioned previously, infants with low birth weight caused by prematurity or maternal smoking are at an increased risk of SIDS. Also at increased risk are those infants who are small-for-gestational-age. United Kingdom figures for 1998 show that the SIDS rate for infants in the 1500—2000g birth weight range is nearly nine times higher than infants of birth weight greater than 3500g (Office for National Statistics, UK, 1998). Interestingly, SIDS infants are generally smaller than control infants, in both weight and length. They have also been shown to have a slower postnatal weight increase as compared to controls, even if initially of normal birth weight (Fujisaki et al., 1990; Williams et al., 1990; Blair et al., 2000).

Postnatal Risk Factors

A multitude of studies have shown that prone sleeping leads to an increased risk of SIDS. The first person to report this relationship, and recommend that infants should not sleep in this position, was Abramson in 1944. This association was generally ignored for around 40 years, until the late 1980s when 'back-to-sleep' campaigns commenced. The public health campaigns promoting a non-prone sleep position have been credited with a drop in SIDS rates of around one half, and up to 70% in certain countries (for example Sweden) (Alm et al., 2001; Dwyer & Ponsonby, 1995; Guntheroth, 1995). Advice on infant sleep position has been directed, in the main part, on not placing infants prone. Yet there is some evidence showing that placing infants to sleep on their sides, while being safer than the prone position, is associated with an increased risk of SIDS (Oyen et al., 1997; Fleming et al., 1996a). Reports from the New Zealand Cot Death Study showed, after adjusting for other confounding variables, a significantly high risk of SIDS from the prone position compared to the supine position (OR 6.7), but also reported a significant increase in SIDS risk associated with the side sleeping position (OR 2.0) (Mitchell et al., 1992a).

Breast feeding has been found to be a significantly protective factor in many recent studies (Guntheroth, 1995). For example, the New Zealand Cot Death Study (Ford et al., 1993) found that SIDS risk was significantly reduced in breast fed infants, after adjustment for potential confounding demographic, maternal and infant factors. Similarly, after controlling possible confounders, Harper & Hoffman's (1988) study found a significant protection against SIDS from breastfeeding. Conversely, Kraus et al. (1989) showed that lower rates of breast feeding for SIDS infants were generally explained by controlling for prematurity and socio-economic status.

There is some evidence that sleeping an infant on a soft surface, using duvets, and overheating of the infant can lead to increased risk of SIDS (Guntheroth, 1995; Sullivan & Barlow, 2001). Bed sharing with adults, historically regarded as the cause of sudden and unexpected deaths continues to be controversial; bed sharing may lead to increased breastfeeding rates, a potential protective behaviour, however it may also lead to an increase in exposure to passive smoking, softer beds, duvets and pillows, thus increasing SIDS risk. Blair et al. (1999) found no evidence of increased SIDS risk for infants when bed sharing with non-smoking adults.

Although SIDS rates are generally lower in communities where the common infant care practice is to sleep infants in the same room as parents or siblings, there has only been one detailed study into the effects of room sharing on SIDS. The New Zealand Cot Death Study found a protective effect from an infant sharing a room with one or more adults, while room sharing with a child had no effect, and even possibly led to increased SIDS risk (Scragg et al., 1996). The risk of SIDS from prone sleeping was reduced when sharing a room with an adult. Quantifying these results showed that an infant sleeping supine in a room with an adult had the lowest risk of SIDS. In comparison, an infant room sharing with an adult and sleeping prone was only three times more likely to succumb to SIDS, while a prone sleeping infant in a room alone was 17 times more likely to die (Scragg et al., 1996).

Pacifier use is another contradicting risk behaviour; recent studies have shown that pacifier use may lead to a reduced risk of SIDS (Mitchell et al., 1993b; Arnestad et al., 1997; L'Hoir et al., 1998), yet pacifier use has also been associated with decreased duration of breast feeding, thus increasing SIDS risk (Fleming et al., 1996b).

In summary, many factors have been associated with both increased and decreased risk of SIDS. Risk factors are not acting independently, with many identified and hypothesised inter-relationships between variables. Many of the identified risk factors are not unique to SIDS, but are also associated with other causes of infant mortality. A single risk factor that exists across all SIDS cases has yet to be identified. Engelberts (1991) suggested that SIDS are a result of two general types of risk factors: "some [that] increase the vulnerability of a child and some [that] prove too great a stress in vulnerable infants".

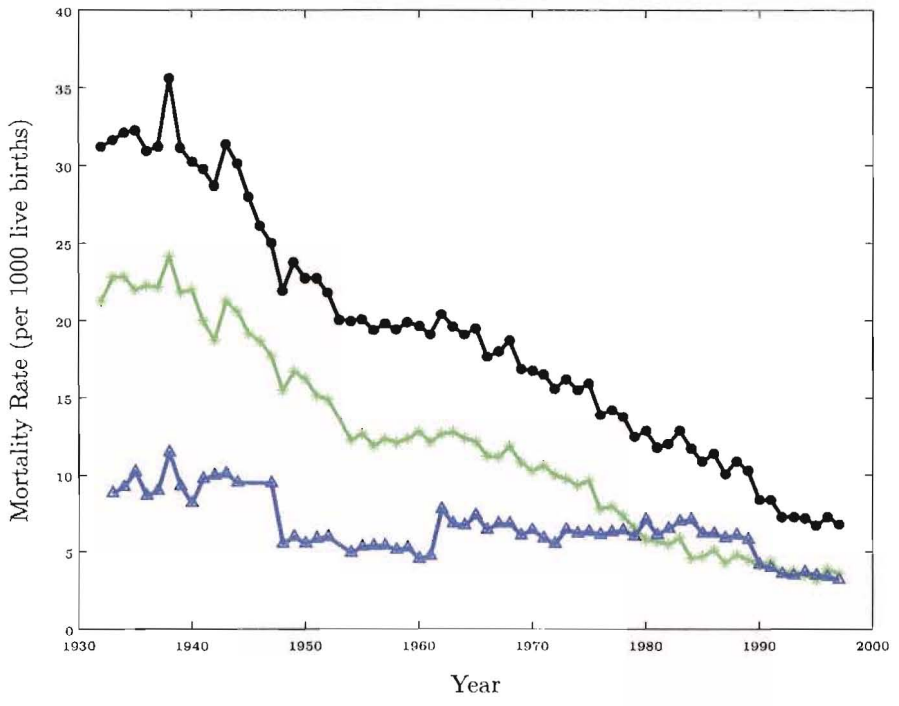


Figure 1.6: New Zealand Infant mortality rates (• = Total infant mortality; + = Neonatal mortality; △ = Postneonatal mortality).

1.6 SIDS in New Zealand

New Zealand's infant mortality rate has been declining steadily for the last 50 years (Figure 1.6). This improvement is primarily a result in dropping neonatal mortality rates, while postneonatal rates remained reasonably stable from the second world war until the commencement of 'back-to-sleep' campaigns. While New Zealand's neonatal mortality ranks as one of the lowest in the world, postneonatal rates remain high compared with other developed countries. This is almost entirely a result of the very high SIDS rate in New Zealand, peaking at 5 SIDS per 1000 live births in 1984 (Fetal and Infant Deaths, 1986). The New Zealand SIDS rate, as with the postneonatal mortality rate, is one of the worst compared to other industrialised countries. SIDS is classified as the principal cause of postneonatal deaths in New Zealand; SIDS caused 53% of all postneonatal deaths in 1986, dropping marginally to 43% in 1998 (Fetal and Infant Deaths, 1986; 1998). The direct relationship between SIDS and postneonatal mortality is illustrated in Figure 1.7. This figure presents SIDS and postneonatal death rates in New Zealand for the twelve year period commencing in 1986.

In an attempt to redress the unusually high SIDS rate, the New Zealand Cot Death Study was developed. This was a large, multicentre case-control study which was completed over three years, commencing in 1987 (Mitchell et al., 1989). One of the main aims of the study was to identify risk factors which were related to an increased risk of New Zealand infants dying of SIDS. The study identified many variables associated with either an increased risk of SIDS or providing a protective effect from SIDS, and resulted in a nationwide cot death prevention programme promoting four modifiable risk factors: prone sleeping position, maternal smoking, formula feeding and bed sharing with infants (Mitchell et al., 1992c). New Zealand was one of the first countries in the world to formally launch a 'back-to-sleep' campaign, and as with other countries, the

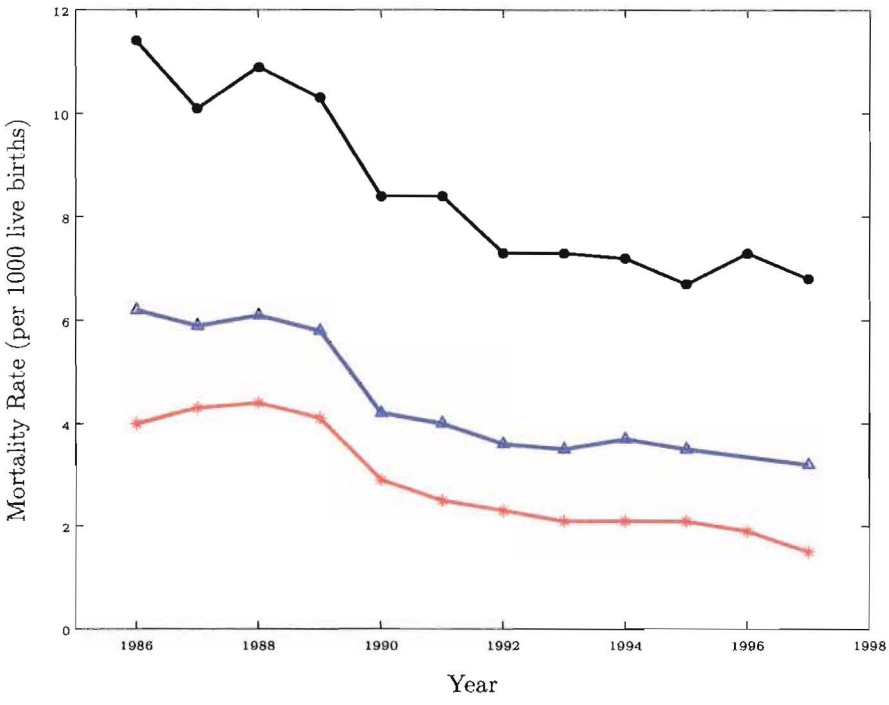


Figure 1.7: SIDS, postneonatal and total infant mortality rates in New Zealand, 1986 to 1997 (• = Total infant mortality; △ = Postneonatal mortality; * = SIDS).

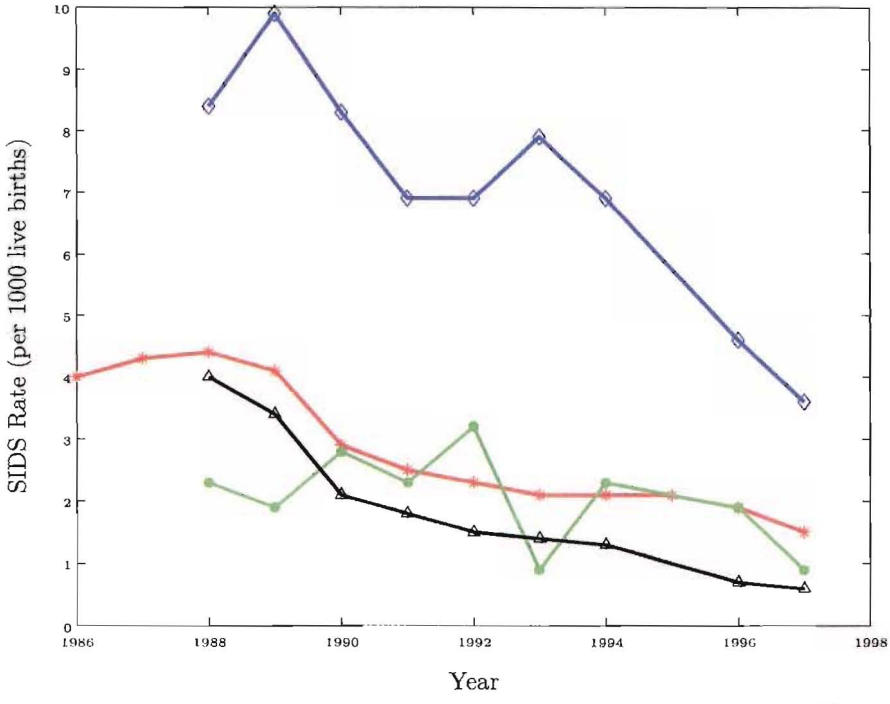


Figure 1.8: SIDS mortality, by ethnicity, in New Zealand, 1986 to 1997 (* = Total SIDS; ◇ = Maori SIDS; • = Pacific Island SIDS; △ = Other SIDS).

campaign resulted in a dramatic reduction in SIDS numbers. In 1998 the New Zealand SIDS rate was 1.2 per 1000 live births, a 76% reduction from the peak in 1984. The largest drop occurred from 1989 (4.1 SIDS per 1000 live births) to 1990 (2.9 SIDS per 1000 live births) (Fetal and Infant Deaths, 1986; 1998).

Maori SIDS rates remain very high post 'back-to-sleep' campaigns and are more than double those of other ethnic groups, including Pacific Islanders. The largest reduction in SIDS rates occurred in the non-Maori, non-Pacific Islander ('other') ethnic grouping with an 82% drop from 1989 to 1996. In comparison the Maori rate dropped by 57% and the Pacific Islander rate by 52% over the same time period. This is illustrated in Figure 1.8. The reasons for the ethnic differences remain unknown, although by analysing data from the New Zealand Cot Death Study, Mitchell et al. (1993a) found a higher prevalence of major risk factors in the Maori population, compared with non-Maori, leading to an increased risk of SIDS for Maori infants.

1.7 SIDS in Canterbury

This study examines the incidence of SIDS in Canterbury, New Zealand, for the period 1968—1999. The SIDS data was collected retrospectively by the Community Paediatric Unit, who examined the records of all postneonatal deaths occurring in the region, over the period of the study. Over the period of the study, there were 658 infant deaths classified as SIDS. This equates to an average of 20.6 SIDS per year, with 3.4 SIDS on average in summer and 7.5 SIDS in winter.

The time frame of this study incorporates the change in the disease classification of SIDS. In 1979, the ninth revision of the International Classifications of Diseases (ICD) (World Health Organisation, 1977), was introduced in New Zealand for the collection of death statistics. This included the new category, specifically for cot death or sudden infant death syndrome: ICD code 798.0. Prior to 1979, coronors' pathologists were using the eighth revision of the ICD code (World Health Organisation, 1965). Therefore deaths which nowadays would be defined as SIDS, were classified into other disease categories, often either a respiratory or viral coding (Ford, 1986).

To ensure a consistent cause of death over the 32 years of the study, three categories were defined as follows: inevitable deaths, preventable deaths, and cot deaths (SIDS). The definition of a SIDS in this study is therefore:

> *Deaths in infants that had been specifically labelled as cot deaths or the sudden infant death syndrome (SIDS), and deaths of ill defined respiratory cause or unknown cause that nowadays would be given the SIDS label* (Ford et al., 1990).

In the 1980s, Canterbury had one of the highest SIDS rates in the western world (Ford, 1986). Figure 1.9 presents the annual profile of SIDS per 1000 live births over the years of this study (1968—1999)[†]. The figure shows SIDS rates increasing steadily to a peak in the eighties. Rates then drop sharply round 1989—1991. Also shown on Figure 1.9 is the estimated prone sleeping rate for the years when this data was available. The changes in the SIDS profile appear to mirror the changes in the prone sleeping profile. The sharp decline in both SIDS and prone sleeping rates corresponds to the time when media publicity campaigns commenced, promoting back or side sleeping, breast feeding, a smoke-free environment, and avoiding bed sharing (Mitchell, 1993). These visual trends highlighted in Figure 1.9 will be examined statistically in Chapter 2 using three novel approaches to the change point problem.

[†]The prone sleeping data was sourced from an unpublised report (1998) by Prof R.P.K. Ford (Community Paediatric Unit, Christchurch), which itself was a compilation of multiple New Zealand studies.

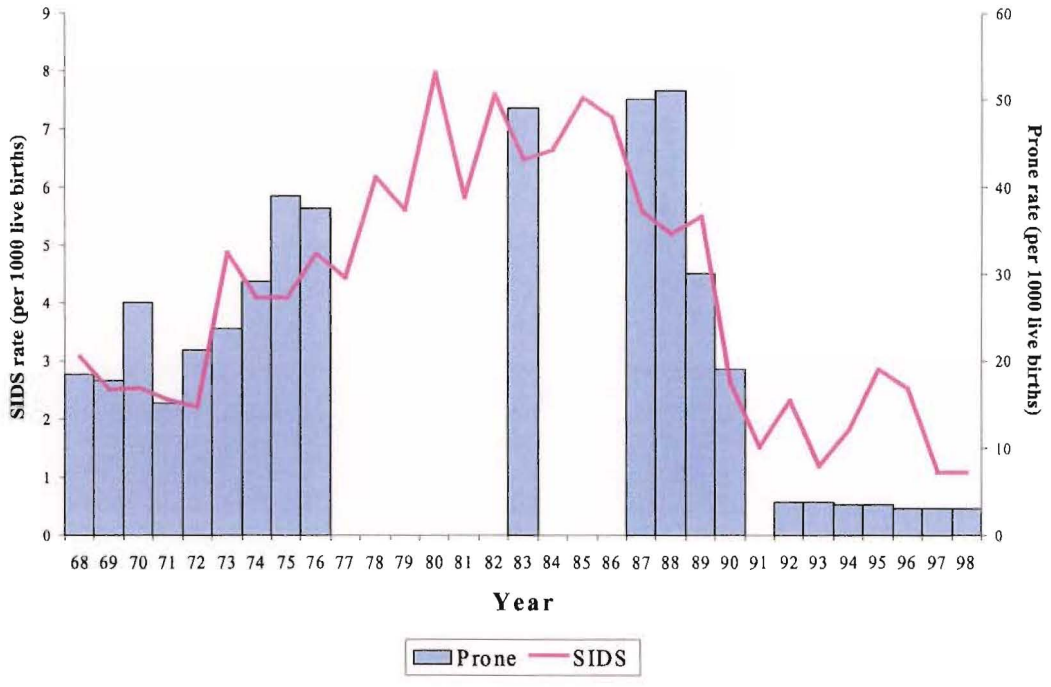


Figure 1.9: SIDS and prone sleeping rates in Canterbury, New Zealand, 1968—1998.

Previous studies using this SIDS profile have included two climate based analyses. Firstly, Schluter et al. (1998) examined daily SIDS counts from 1968—1989 in relation to both long and short term temperature associations. The second study by Macey et al. (2000) investigated the relationships between the incidence of SIDS and wind direction and speed, over and above the temperature based relationships identified by Schluter et al. (1998). Macey et al. (2000) examined daily SIDS counts, over the years 1968—1997. Both these studies, and their corresponding findings, are discussed further in Chapter 1.8.

1.8 Definitions and Construction of Climate Variables

Research has, for a long time, looked to climatic variables as a contributor to sudden infant death syndrome (SIDS) risk. Many studies have related SIDS to various meteorological measures throughout the world, yet the only consistently found relationship is between SIDS and seasonality. As discussed in Section 1.4.1, there is a seasonal cycle within SIDS series, with more infants dying in winter and less in summer (see for example Douglas et al. (1998) or Mitchell et al. (1999)). This trend has been seen in such countries as Britain, Australia, U.S. and Sweden. There is no consistency in seasonality measures across studies, with methods ranging from month indicator variables (Campbell, 1989), monthly temperature averages (McGlashan & Grice, 1983) and retrospective moving averages of temperature (Schluter et al., 1998), to sinusoids and their second harmonics (Douglas et al., 1998; Campbell et al., 2001).

Temperature is the most common meteorological variable examined for its potential effects on SIDS. Though long-term temperature has consistently been related to SIDS through its use as a proxy for season, conflicting results have been reported for a short-term temperature SIDS relationship (as measured by lags of up to nine days). For example, Campbell (1989) examined daily SIDS numbers in New South Wales,

Australia, in relation to mean daily temperature measured in Sydney. Using Poisson regression, he found that a fall in temperature increased the incidence of SIDS six days later, over and above the usual seasonal fluctuations of SIDS (measured by a monthly indicator variable). He concluded that it was this deviation in temperature, from the monthly average, that resulted in increased SIDS incidence, not the temperature itself.

A similar inverse lagged relationship between temperature and SIDS was identified in later work by Campbell (1994) which related daily SIDS in England and Wales to maximum and minimum daily London temperatures. Campbell used an autoregressive Poisson regression technique, to show that SIDS risk increased two to five days after a drop in temperature, over and above the seasonality relationship measured by sinusoids.

Mitchell et al. (1992b) examined the relationship between SIDS and daily minimum temperature in Auckland and Christchurch, New Zealand. After accounting for seasonality (using a sinusoid), Mitchell et al found a significant association of SIDS with minimum temperature four and five days before a SIDS death. They found that on average, the days before a death were colder than other days.

McGlashan & Grice (1983) examined the relationship between daily SIDS numbers and daily minimum temperature (measured in Hobart), alongside monthly minimum temperature, and the difference between daily minimum temperatures on consecutive days in Tasmania. They found that the SIDS numbers decreased in months with higher minimum temperatures (highlighting the seasonality effect); that more SIDS occurred on days where the minimum temperature was less than 8°C ('cold' days); and that sharp changes in minimum temperature from the preceding day ($> \pm 3°$) was associated with fewer SIDS. In McGlashan & Grice (1983) there was no control for seasonality with these last two threshold-type effects, and therefore these effects may in fact be acting as season proxies.

Schluter et al. (1998) examined a subset of the Canterbury SIDS data (1968—1989) analysed in this thesis. In Schluter's study, daily measures of minimum, mean and maximum temperatures, temperature variations (range and standard deviation), 'between day' temperature measures (temperature lags), and seasonality measured via a 31-day retrospective moving average of temperature were examined. Ordinal logistic regression was used to confirm the annual cyclicity in SIDS, and it was found that days with a warmer minimum temperature, or with little temperature change, were significantly related to a higher SIDS incidence, over and above the seasonal SIDS pattern. No significant temperature lag effect was found. This work was unusual within the SIDS literature due to the localisation of the recorded SIDS deaths around the point of observation of the temperature; most SIDS deaths occurred within a 20km radius of the meteorological station.

There have only been a small number of studies relating SIDS to meteorological variables other than temperature. Those that have include Macey et al. (2000) who also looked at a subset of the Canterbury SIDS data set (1968—1997). Their objective was to examine SIDS incidence in relation to wind on a daily level, focussing on the föhn, or north-west wind. Macey et al. (2000) examined a polytomous wind direction variable, including 'calm' and 'mixed' categories, two daily wind direction variation variables, as well as mean, maximum and standard deviation measures of daily wind speed and corresponding daily lags. The SIDS data was split into exploratory and confirmatory data sets and analysed using Poisson regression. The temperature model found by Schluter et al. (1998) was confirmed by Macey et al., and significant relationships between SIDS and wind were found using the exploratory data. Variables identified as significantly associated with increased SIDS risk by Macey et al. included northerly winds on the day of death, and southerly winds occurring three days before death, after controlling for temperature effects. Validation of these wind effects with the second data set failed to find any significant associations and it was concluded that SIDS incidence in Canterbury is not significantly influenced by wind. The validation process was performed on post 'back-to-sleep' campaign Canterbury data, where the SIDS numbers had dropped

dramatically (1.89 SIDS per 1000 live births per year as compared to 5.10 SIDS per 1000 live births per year pre 1990) leaving insufficient numbers for analysis. This may have been the cause of the lack of significant results. In fact, in Chapter 2, Section 2.6 it was shown that SIDS deaths in 1990—1999 occur at random time intervals. Therefore, the lack of significant wind effects in the validation process is not surprising.

Campbell et al. (2001) investigated atmospheric pressure and SIDS in Chicago. Using autoregressive Poisson models, they found a weak positive link with mean daily atmospheric pressure, over and above seasonal effects (measured by sinusoids and their second harmonics), and mean daily temperature. They examined, but failed to confirm, a one or two day lag effect with atmospheric pressure.

In 1979, Deacon et al. introduced an insolation index variable, based on solar radiation and temperature in winter. This index was found to be positively correlated to SIDS incidence in Australia and Britain. A second index was further developed by Deacon & Williams (1982), which incorporated rainfall along with insolation and temperature. Again, this index was positively associated with British and Australian SIDS incidence. Deacon & Williams (1982) suggested that it may be the combination of dampness and low temperatures that is crucial in SIDS aetiology.

Nelson & Taylor (1988) performed a regional analysis of SIDS rates in New Zealand relating to SIDS climatic, ethnic and socio-economic factors. They examined mean daily temperature and vapour pressure, alongside Deacon's insolation and cold-wet weather indices. Their results showed significant correlations between regional SIDS incidence and temperature, pressure and the insolation index, but failed to show a significant relationship between the cold-wet index and SIDS. They suggest that, in New Zealand, the effect of rainfall on SIDS is negligible compared to temperature, pressure and solar radiation.

A recent study by Lipfert et al. (2000) looked at the effect of air quality on SIDS incidence in the U.S.. Ecological and environmental effects on SIDS were investigated, using birth and death records coupled with county-level air quality data, via a cross-sectional analysis. In their logistic regression analyses, a dummy month variable was used to control for seasonality and, after accounting for this and birth, ethnic, and maternal factors, their results showed a strong negative spatial relationship with SO_4^{2-} aerosol (often the primary component of fine particles ($PM_{2.5}$)) and a positive relationship with PM_{10}. Lipfert et al. (2000) speculate that indoor pollution from domestic heating, such as wood fires, may in fact be an important environmental risk factor associated with SIDS.

Another recent study by George et al. (2001) examined pollution in relation to SIDS, specifically looking at nitrate levels in drinking and ground water in Sweden. Monthly SIDS rates were examined for geographical trends using a space-time analysis, and Poisson regression was performed to examine how SIDS incidence related to monthly and quarterly seasonality indicators. Both geographical and seasonality trends were significant. Spearman rank correlations were used to show a positive correlation between SIDS and maximally recorded concentrations of nitrate in drinking water, but did not account for the previously identified seasonal, or geographical, relationships.

Two studies have examined SIDS incidence comprehensively in relation to multiple meteorological variables and air pollution: Auliciems & Barnes (1987) examined SIDS incidence over a 15 year time frame in Brisbane, Australia, while Knöbel et al. (1995) looked at SIDS over an 11 year period in Taiwan.

Auliciems & Barnes (1987) used a time-track analysis to examine the weather conditions over the 24-hour period before each recorded SIDS in their study. They used a presence/absence indicator as an approximate index of particulate pollution, alongside hydrometeors (precipitation in any form), wind speed, air temperature, total cloud cover and visibility measures, for each of the 29 days preceding and 20 days following a SIDS death, to examine the time-track patterns for any associations with SIDS. They found, unexpectedly, pollution levels dropped over the days prior to SIDS. They also found strong evidence of atypical time traces in the region of SIDS for visibility: in both summer and winter SIDS cases, mean visibility increased preceding death. No other weather associations with SIDS incidence were found. They hypothesised that these

findings of higher SIDS incidence after clear weather may be a result of either indirect effects of light, or infants having an increased exposure to the outdoors and/or clear weather influencing changes in parental patterns.

Knöbel et al. (1995) used Poisson regression analysis to relate daily SIDS rates in Taiwan to daily average values of visibility (an optimetrical measure of air pollution), temperature, air pressure, sunshine duration, rainfall, relative humidity and wind speed. They validated the use of visibility as a proxy for gravimetrical air pollution measures (PM_{10}, PSI, sulphur dioxide and carbon monoxide) by correlating visibility with pollution over 2 years of the study time frame. Significant negative associations were found between the four air pollution measures and visibility. In a combined regression analysis, Knöbel et al. (1995) found that only visibility and temperature were significantly associated with SIDS incidence. Their results showed that SIDS deaths were better predicted by visibility measures nine days prior to SIDS, as compared to measures recorded on the day of death. Their analysis did not include a seasonality variable, though they reported the commonly found seasonal pattern of SIDS.

As with the studies presented by Auliciems & Barnes (1987), and Knöbel et al. (1995), this study aims to undertake a comprehensive analysis of SIDS incidence in relation to multiple meteorological measures in Canterbury. Unlike the study by Auliciems & Barnes, which examines climatic conditions centred around the day of death, regression-type analytical methods are applied to examine the 11688 days of Canterbury SIDS data to build up a profile of 'at-risk' SIDS days. This study extends previous research by examining nine measured climatic variables, including dewpoint, which has not been previously examined in relation to SIDS incidence, alongside wind velocity and wind chill, over an extended 32 year time frame. This is the first complete analysis, undertaken via systematic and methodical processes, to examine the climatic profile of SIDS risk with respect to such a detailed set of recorded weather variables. This study has the potential to result in a daily climate index which, while incorporating a large amount of meteorological information, relates simply to SIDS.

1.8.1 Canterbury Climate Data

The Canterbury climate data used in this study is a detailed data set with the variables recorded as hourly observations over 32 years from 1968—1999 (except for sunshine and solar radiation, which were only available at a daily level) giving nearly 2.5 million weather observations. This fine measurement scale allowed Schluter et al. (1998) and Macey et al. (2000) to create comprehensive day, within day and between day measures for temperature and wind respectively. Similar comprehensive variables covering all the available climatic information are defined, and extended, in this study. This results in approximately 2500 daily climatic measures which will then be examined for potential relationships with the incidence of SIDS.

A hierarchical approach is used to examine the Canterbury SIDS—climate relationship at a daily level, incorporating seasonal effects, day effects, daily lagged effects, between day effects, and within day effects. Principal components are then used to develop a model containing multiple climatic covariates to predict SIDS.

The climate data was recorded by the National Institute of Water and Atmospheric Research (NIWA) at the Christchurch International Airport, over the years, 1968—1999. This airport is judged to encounter weather conditions typical of the Canterbury plains region because of the exposed nature of the site. Figure 1.10 shows the geographical location of the recording site in relation to the national geography of New Zealand. The 20km radius around the Christchurch International Airport is also marked; 90% of SIDS deaths occurred within this circumference. The general Christchurch area has a homogeneous climate, with the exception of the Port Hills to the south. Few SIDS deaths occurred in this isolated microclimate.

The climatic variables available for analysis, including the two derived measures of wind velocity and

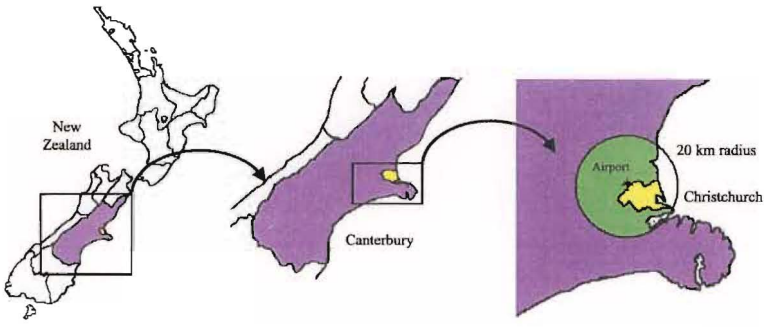


Figure 1.10: Geographical location of Christchurch, within New Zealand. Also illustrated is the location of Christchurch International Airport where the meteorological data were recorded, with the 20km radius where 90% of the SIDS death occurred highlighted.

Variable	Notation
Temperature (°C)	Temp
Wind Direction	WindD
Wind Speed (knots)	WindS
Wind Velocity (knots)	WindV
Wind Chill (°C)	WindC
Humidity (%)	Humid
Pressure (HPa)	Pres
Rainfall (mm/hr)	Rain
Sunshine (hrs)	Sun
Solar Radiation (Mj/m^2)	Rad
Visibility (km)	Vis
Dewpoint (°C)	Dew

Table 1.3: Climatic variables.

wind chill, are listed in Table 1.3. Data was recorded hourly, with the exception of sunshine and solar radiation, where only daily data was available. Thus, for the 32 years, or 11688 days, of the study, there are 2 460 336 climate measures. The following sections give a detailed description of the climatic variables used in this study, including definitions, measurement procedures and daily summary variables. Note: a day is defined as the 24 hour period to midnight.

As shown in Chapter 2, the distribution of SIDS falls into three distinct periods, denoted Period 1 (1968—1972), Period 2 (1973—1989) and Period 3 (1990—1999). It was shown that in Period 3, the SIDS distribution did not contain any underlying trends or seasonal patterns, essentially it was a random series. Therefore, throughout the analysis of the relationship between climate and the incidence of SIDS in Canterbury, Period 1 and Period 2 are examined separately as distinct datasets, with Period 3 excluded from further analysis due to the random nature of the occurrence of SIDS in this period. For completeness, some reference is made to the climate data recorded in Period 3. See Chapter 2 for further details. For added readability, many of the figures discussed in this section are presented in Appendix B.

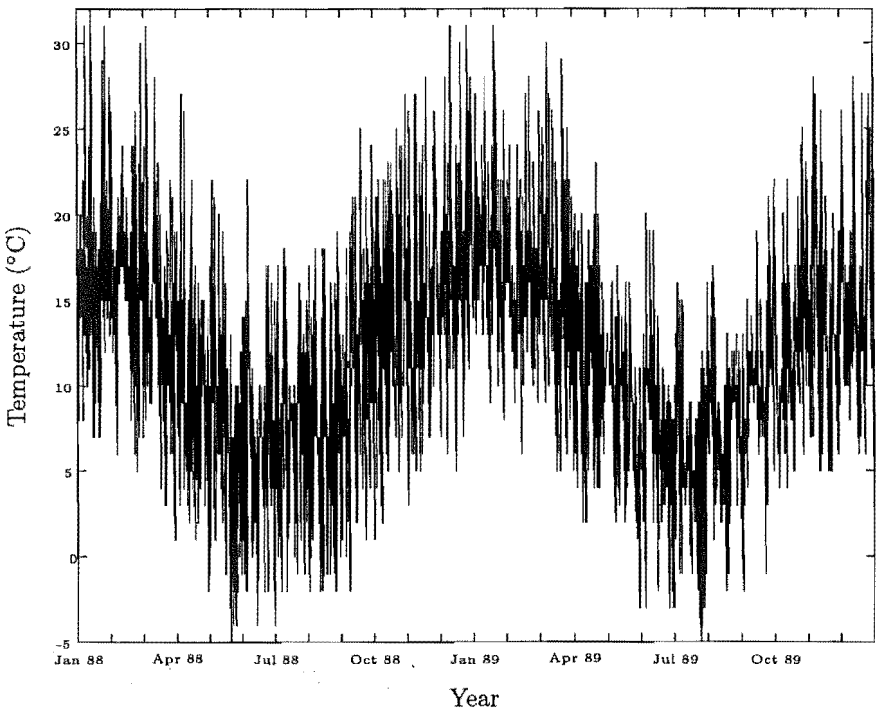


Figure 1.11: Hourly temperature measures for 1988—1989.

1.8.2 Temperature

> ***Temperature:*** *the concentration of sensible heat in a body, measured according to an arbitrary scale such as the Celsius or Kelvin scales* (McIlveen, 1986).

Air temperature is a notable atmospheric characteristic, yet although human bodies are responsive to the ambient air temperature, this is confounded by many other factors. The impression of being hot or cold is dependent upon whether heat is being consumed or preserved to maintain internal body temperatures at an equilibrium of 37°C. This, in turn, is dependent on exposure to different climatic conditions including sunlight, terrestrial radiation, wind precipitation or humidity, personal health, and how recently the last meal was consumed, its size, and temperature (McIlveen, 1986).

Spot readings of temperature were taken hourly using a mercury-in-glass (dry bulb) thermometer. For the period 1968—1970 the temperature was recorded in degrees Farienheight (°F), and consequently converted to degrees Celsius (°C) for a consistent unit of temperature measurement over the period of the study. The conversion was completed by the NIWA prior to the attainment of the data. Temperature was recorded to the nearest degree until 1990, and then, due to updated instrumentation, taken to the nearest half a degree.

Figure 1.11 shows the hourly temperature recordings for the years 1988—1989 (chosen purely for illustration purposes). The overall periodicity of the temperature (a warmer temperature in summer, and cooler temperature in winter; on average 16°C and 6°C respectively) can be seen underlying the more random fluctuations in the temperature series. The maximum hourly temperature recorded over the two years shown in Figure 1.11 was 32°C, with the minimum for the same two years being -5°C. The maximum hourly temperature recorded over the complete study period was 40°C, while the minimum hourly temperature was −6°C.

Daily Variables and Notation

Retrospective lagged variables were created, rather than averaging over periods covering pre- and post-death time spans as it was assumed to be the climatic conditions prior to death which affected the incidence of SIDS, rather than the weather after the event.

Days and Lags

The notation for the day of interest and retrospective lags is:

- **day0** refers to today (that is, the day of interest)
- **day-1** refers to the day before the day of interest
- **day-2** refers to two days before the day of interest
- **day-3** refers to three days before the day of interest
- **day-4** refers to four days before the day of interest
- **day-5** refers to five days before the day of interest
- **day-6** refers to six days before the day of interest
- **day-7** refers to one week before the day of interest
- **day-8** refers to eight days before the day of interest
- **day-14** refers to 14 days before the day of interest (a fortnight prior)

Due to the retrospective nature of the variables to be created, and the unavailability of concomitant hourly data for December 1967, the daily temperature (and other climatic) summaries commence on 15 January 1968.

Day and Within Day Effects

The following summary measures were created from the hourly temperature data for each day and lagged-day as detailed above. Note: *AHC* denotes the absolute hourly change in temperature, which is defined as the absolute difference between any two consecutive hourly temperature measures.

- minimum (**min**)
- mean (**mean**)
- maximum (**max**)
- standard deviation (**std**)
- range (**range**)
- maximum AHC (**maxAHC**)
- mean AHC (**meanAHC**)
- standard deviation AHC (**stdAHC**)

Between Day Effects

For each of the eight day and within day temperature measures described above (**min** to **stdAHC**), sixteen between day effects were created as a description of the difference in temperature between days. The following differences were created, where *day-i-j* gives the difference in temperature measure from i to j days ago (for example day-7-8 gives the difference in temperature measure from seven days before the day of interest to eight days before the day of interest).

- **day0-1**
- **day-1-2**
- **day-2-3**
- **day-3-4**
- **day-4-5**
- **day-5-6**
- **day-6-7**
- **day-7-8**
- **day0-2**
- **day0-3**
- **day0-4**
- **day0-5**
- **day0-6**
- **day0-7**
- **day0-8**
- **day0-14**

Average Day Effects

The average temperature over the i days before the day of interest was created for the following periods:

- day-1 and day-2 (**meanday-1to-2**)
- day-1, day-2 and day-3 (**meanday-1to-3**)
- day-1 to day-4 (**meanday-1to-4**)
- day-1 to day-5 (**meanday-1to-5**)
- day-1 to day-6 (**meanday-1to-6**)
- day-1 to day-7 (**meanday-1to-7**)
- day-1 to day-8 (**meanday-1to-8**)
- day-1 to day-14 (**meanday-1to-14**)

Between Average Day Effects

These effects are the difference between the mean temperature on the day of interest ($day0$) and the average temperature over the i days prior to the day of interest (the average day effects described above). A between average day effect that is greater than zero implies that the temperature is warmer on the day of interest than it was over the i days prior, while conversely, a between average day effect less than zero shows that the temperature on the day of interest is cooler than it has been on average previously. The following is used to notate these effects:

- **diffday0-2 = meanday0 − meanday-1to-2**
- **diffday0-3 = meanday0 − meanday-1to-3**
- **diffday0-4 = meanday0 − meanday-1to-4**
- **diffday0-5 = meanday0 − meanday-1to-5**
- **diffday0-6 = meanday0 − meanday-1to-6**
- **diffday0-7 = meanday0 − meanday-1to-7**
- **diffday0-8 = meanday0 − meanday-1to-8**
- **diffday0-14 = meanday0 − meanday-1to-14**

This gives a total of 224 daily summary temperature variables created from the original hourly measures. Full notation for the temperature variables is:

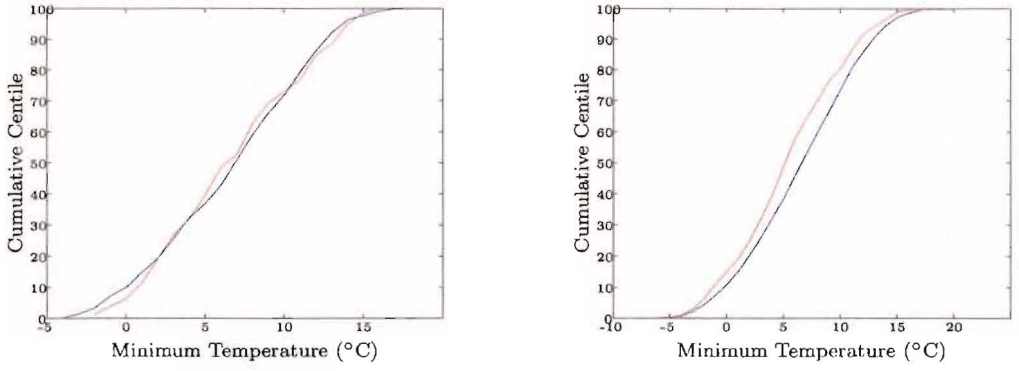
$$Temp_{(summary-measure)(day-span)}. \tag{1.1}$$

For example:

- $Temp_{meanday-5}$ is the mean temperature five days prior to the day of interest.
- $Temp_{stdAHCday-6-7}$ is the difference in standard deviation of the absolute hourly change in daily temperature between six and seven days before the day of interest.
- $Temp_{meanday-1to-7}$ is the mean temperature over the seven days prior to the day of interest.
- $Temp_{diffday0-7}$ is the difference between the temperature on the day of interest ($day0$) and the average temperature over the previous week (seven days).

This notation format will be used for describing the variables created for the other climatic measures used in this study, with $Temp$ being replaced by the variable of interest.

Daily temperature summaries were calculated from the hourly recordings, as detailed above. There were approximately 0.05% missing data over the 32 years, with no one day missing all 24 hourly measures.



(a) Period 1 (b) Period 2

Figure 1.12: Cumulative centile distribution for minimum daily temperature (— SIDS days, — non-SIDS days).

The missing data was ignored. The gross seasonal fluctuations seen in the hourly temperature data were replicated as a cyclic pattern within the minimum, mean and maximum temperature series. The remaining summary temperature variables showed no obvious periodicities or non-random patterns (not pictured).

The cumulative centile distribution of minimum daily temperature, categorised into days with no SIDS occurring and days where at least one SIDS has occurred, is presented in Figure 1.12. In Period 2 (Figure 1.12 (b)) there is a clear differential between SIDS and non-SIDS days: the temperature is lower on days where at least one SIDS has occurred, as compared to days with no SIDS deaths. This suggests a potential relationship between SIDS occurrence and temperature in Period 2, although no seasonal effects have been accounted for in the information depicted in Figure 1.12. For Period 1 (Figure 1.12 (a)) there is no clear distinction between the cumulative centile distribution of minimum daily temperature on days with no SIDS occurring and days where at least one SIDS has occurred, with the distributions overlapping over the range of temperature.

1.8.3 Wind Direction

> ***Wind Direction:*** *the direction* ***from*** *which the air comes*
> (McIlveen, 1986).

Alongside temperature, wind has long been associated with ill health. The warm, dry alpine winds, or föhn winds, have been related to cardiovascular problems, migraines and general increased irritability (Verhoef et al., 1995; Miric & Rumboldt, 1993; Piorecky et al., 1997).

Wind direction was measured hourly as the average direction over the previous ten minutes (to the nearest 10°) using a wind vane with an electromechanical attachment to record the vane direction. The wind vane was upgraded at the beginning of 1993 by NIWA, with the new equipment being more responsive to lighter winds. When the wind speed was zero, the wind direction was recorded as zero, corresponding to a measurement of no wind. Recordings of 990 correspond to hours when the wind direction was variable or mixed. Following the work of Macey et al. (2000), the direction in degrees was resolved into ten wind direction categories as detailed in Table 1.4. The mixed category corresponds to days where no one wind direction predominated, while the direction categories correspond to days when the wind prevailed in one

Direction (Category)	Direction (Degrees)
1. Calm (C)	—
2. North (N)	337.5 — 22.5
3. North-East (NE)	22.5 — 67.5
4. East (E)	67.5 — 112.5
5. South-East (SE)	112.5 — 157.5
6. South (S)	157.5 — 202.5
7. South-West (SW)	202.5 — 247.5
8. West (W)	247.5 — 292.5
9. North-West (NW)	292.5 — 337.7
10. Mixed (M)	—

Table 1.4: Wind Direction categories.

of the eight defined directions. Calm wind direction days are days where there was predominately no wind. In accordance with the absolute hourly change (AHC) defined for temperature, three further daily wind direction variables were created: maximum AHC, mean AHC and standard deviation of AHC (measured in degrees). For hours where the wind direction was either mixed or calm, the absolute hourly change was treated as missing, and ignored. This gave a total of 13.71% hours of missing data for measures of absolute hourly change in wind direction. As with temperature, these variables were calculated for $day0$, and lags up to $day - 14$.

Figures B.1 and B.2 show the predominant seasonal wind direction ($WindD$) distribution in Period 1 (1968—1972) for days where no SIDS occurred and days where at least one SIDS occurred respectively. The same information for Period 2 (1973—1989) is given in Figures B.3 and B.4. The distribution of wind direction categories is very similar across both periods for days where no SIDS occurred. In contrast, with a much smaller number of SIDS in Period 1 than Period 2, the wind direction distribution for Period 1 appears sparse, although the similarities across Period 1 and Period 2 occur in both winter and spring. Figures B.1 to B.4 highlight the lack of south-easterly winds in Canterbury (0.87% of hourly recordings over the complete study period), and the predominance of easterly winds in summer and spring. In winter, calm wind conditions are the most common, while the south-westerly winds, often associated with very bitter conditions, also feature strongly in winter and autumn. There appears little difference in the wind direction distribution between days where at least one SIDS occurred as compared to days with no SIDS. This is not surprising, as by itself, wind direction is not expected to supply any useful information with respect to SIDS incidence. For instance, south-westerly winds in Canterbury are associated with cold-wet conditions, near-calm high pollution nights and beautiful summer days on a dying south-westerly episode. Conversely, north-westerly winds are associated with both near-calm high pollution nights and extremely hot föhn conditions[†].

1.8.4 Wind Speed

> ***Wind:*** *air motion relative to the Earth's surface, usually the horizontal component* (McIlveen, 1986).

A munro anemometer, or anemograph, at 10m was used to measure the horizontal component of wind speed. The hourly measures were derived by taking the mean speed for the ten minutes to the hour and

[†]Personal communication, Allan Ryan, March 2002.

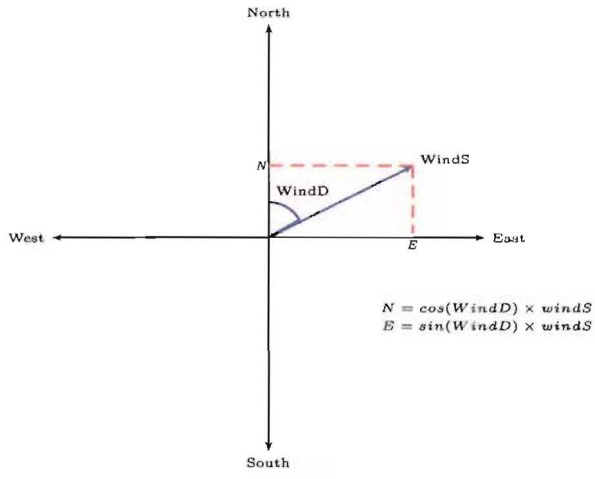


Figure 1.13: Resolving wind speed (WindS) into north and east directions.

recorded to the nearest knot. Over the period of the study, the wind speed reached storm level (above 48 knots) on 5 days, gale force (34 - 48 knots) on a further 39 days and was classified as strong (22 - 33 knots) for 1769 days (15% of the study). Hourly readings are consistent from 1968 - 1999 and not affected by the instrumentation upgrade in 1993. Figure B.5 illustrates hourly wind speed recordings for 1988 and 1989. There appears to be a slight seasonal trend within the series, with stronger winds in the warmer seasons (on average 9.48 knots in summer and 6.60 knots in winter). Random peaks within the series almost swamp this periodic effect.

Daily summary variables of minimum, mean, maximum, standard deviation, range and AHC measures were calculated for wind speed ($WindS$) as detailed previously for temperature. There were approximately 0.04% of hourly data missing over the 32 year series, with no one day missing all 24 hours of data. The missing data was ignored. The small seasonal fluctuations seen in the hourly data were replicated as a cyclic pattern within the mean, maximum and mean absolute hourly change wind speed series. The remaining summary wind speed variables showed no obvious periodicities or non-random patterns, with the exception of the necessary censoring of minimum daily wind speed at zero knots (not illustrated). Between, average and between average day effects were also calculated for wind speed.

The cumulative centile distribution of mean daily wind speed, categorised into days where no SIDS occurred and days where at least one SIDS occurred is presented in Figure B.6. There appears to be little difference in the cumulative distributions of the two categories, possibly indicating little association between the incidence of SIDS and wind speed (without accounting for season) in Period 1 (Figure B.6 (a)) or Period 2 (Figure B.6 (b)).

1.8.5 Wind Velocity

> ***Wind Velocity***: *the vector specified by wind speed and wind direction.*

The variable created by resolving wind speed into its component directions gives a vector that has both magnitude and direction. This wind velocity variable contains more information than its component parts (wind direction and wind speed) as it allows for a distinction between light and strong winds in a given direction.

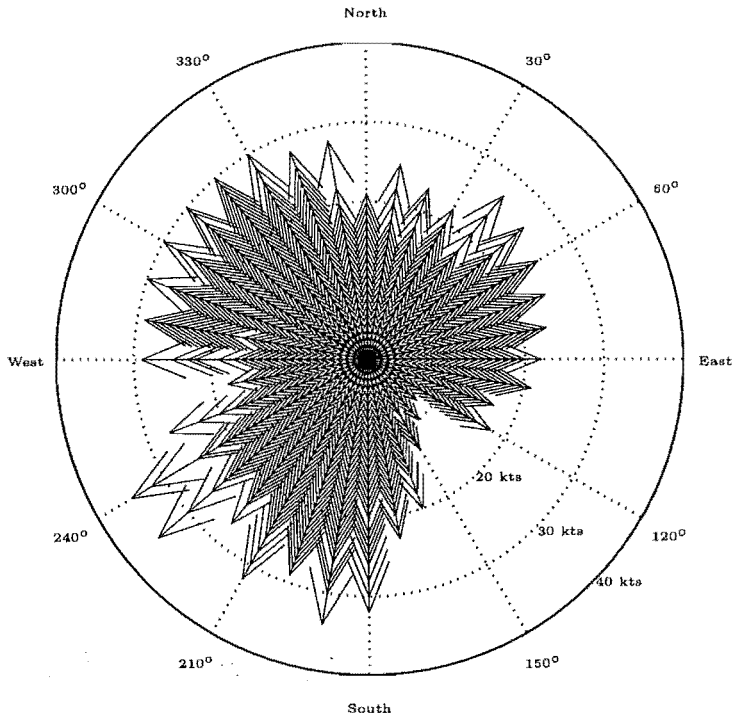


Figure 1.14: Hourly Wind Velocity measures for 1988—1989.

The creation of the wind velocity variable involved a two step process. Initially two vectors were created from the hourly wind direction and wind speed data, giving the speed from the north and east directions. This process is shown in Figure 1.13, which illustrates the trigonometric relationships between wind direction, wind speed and the component directions. A negative speed in the northerly direction corresponds to that speed in the southerly direction, and similarly with negative easterly speeds corresponding to westerly measures. Using these relationships four further hourly variables were created: the wind speed in each of the four primary directions, north, east, south and west (given notation $North$, $East$, $South$, $West$). This process essentially describes wind velocity as a maximal positive basis, a method which has been shown to be useful in a wide range of applications including numerical optimisation (Price et al., 2002). The four wind velocity variables are implicitly linked; in any hour, there will be velocity measures in, at most, two directions once resolved into their component parts. Missing values of wind direction or wind speed were ignored, and hours where there was no wind or no predominate wind direction were assigned wind velocity values of zero in each of the four component directions.

Figure 1.14 presents the hourly wind velocity series for the years 1988—1989. The strongest winds are those coming from the south-west, followed closely by those coming from the north-west. As first seen in Section 1.8.3, winds coming from the south-east are unusual in Canterbury.

From the four direction variables, which give the wind speed in that direction for each hour, daily summary variables were created. These include daily mean, maximum, minimum, range, standard deviation and absolute hourly change measures for $day0$, lagged days, and between day effects, as detailed previously. Of the four directions, the easterly component of wind velocity was the only direction with obvious seasonality

within the daily summary measures. The maximum wind velocity in the eastern direction had the clearest cyclic periodicity, with the strongest winds occurring in summer. The mean, standard deviation and range of easterly wind velocity also exhibited seasonal tendencies (not shown). The south and west wind velocity components gave the strongest wind velocity measures, whilst the mean AHC of wind velocity was the lowest coming from the east (not shown).

Figure B.7 presents the cumulative centile distributions for the daily mean of the four wind velocity directions in Period 1, grouped into days where no SIDS occurred and days where at least one SIDS occurred. The northern component of wind velocity (Figure B.7 (a)) is the only wind velocity direction where there appears be a difference in the cumulative distributions of SIDS or non-SIDS days; the northern wind velocity was stronger on days with at least one SIDS death. The cumulative centile distributions of the four wind velocity directions in Period 2 are presented in Figure B.8. No obvious differences in the cumulative distributions of SIDS or non-SIDS days are evident for wind velocity from both the northerly and southerly directions (Figure B.8 (a) and (b)). In contrast, wind velocity from the east was stronger on days where no SIDS occurred (Figure B.8 (c)) and the westerly wind velocity was stronger on days with at least one SIDS death (Figure B.8 (d)). This highlights a potential relationship between wind velocity and SIDS, especially from the east or west directions, although no seasonal effects were incorporated into this graphical examination of the wind velocity data.

1.8.6 Wind Chill

> ***Wind Chill:*** *the chilling effect of the wind in combination with a low temperature* (Environment Canada, 2002).

Humans do not directly sense the air temperature. It is actually the temperature of their skin that they are sensing when they feel cold. Humans' perception of temperature differs when there is wind, as skin temperature is lowered in its presence. It is this sensation that the wind chill index aims to quantify. In the quantifying process, wind chill is expressed on a temperature scale, yet it is not a temperature, rather it is expressing a human sensation.

Environment Canada have developed a new wind chill index, which is based on a model of how quickly heat is lost from the human face (Environment Canada, 2002; Santee, 2002). The formulation is

$$WindC = 13.12 + 0.6215Temp - 11.37WindS^{0.16} + 0.3965Temp \times WindS^{0.16}, \tag{1.2}$$

where $WindC$ is the wind chill index, based on the Celsius temperature scale; $Temp$ is the air temperature in degrees Celsius (°C), as previously defined in Section 1.8.2; and $WindS$ is the wind speed at 10 metres, in kilometres per hour, as previously defined in Section 1.8.4, and converted from knots.

Wind chill becomes meaningless when the temperature is too warm, or when there is little wind. Therefore, it is only defined for temperatures less than 5°C, and for wind speeds greater than 5 km/hr (2.9 $knots$). Figure 1.15 presents daily wind chill measures for 1988 and 1989, and shows that wind chill, with the above restrictions, is only defined for the colder months of each year. The wind chill index is mainly used to inform people of the risk of injury they face when it is cold, therefore it is non-sensical to have an index over summer months. Environment Canada (Environment Canada, 2002) present 'wind chill hazards', which allow interpretation of this index for various wind chill levels. Their information ranges from a wind chill index of −10°C, giving a slight increase in discomfort, to a level of −60°C or colder, where exposed skin may freeze in less than two minutes. With the Canterbury data, the minimum wind chill was −9.62°C, much higher than index readings that can occur in Canada, yet wind chill may still be a concern in relation to SIDS incidence.

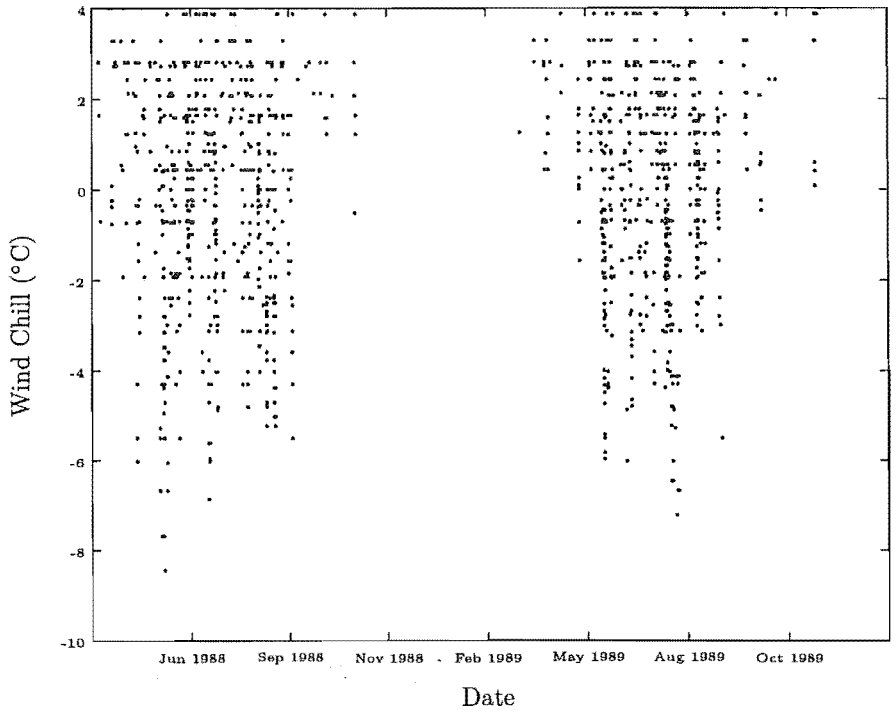


Figure 1.15: Hourly wind chill measures for 1988—1989.

With the restrictions placed on the wind chill calculations, there were only 22600 hours with a wind chill measure, out of a possible 280 512 hours. This leads to summary measures being defined for only 3945 days, or 33.8% of days, over the 32 year time frame of the study. Daily summary variables for minimum, mean, maximum, standard deviation, range maximum absolute hourly change (AHC), mean AHC and standard deviation of AHC were calculated from the hourly wind chill data. The wind chill was considered missing, and ignored when undefined. Due to the smaller dataset, only daily summary variables and lags were calculated.

The cumulative centile distribution of mean daily wind chill (for days where wind chill was defined), categorised into SIDS and non-SIDS days, is presented in Figure B.9. This clearly shows that, for both Period 1 (Figure B.9 (a)) and Period 2 (Figure B.9 (b)), wind chill was greater on days with no SIDS deaths as compared to days where at least one SIDS occurred. This may be due to the direct dependency on temperature in the formulation of wind chill, as temperature exhibited a similar trend with respect to SIDS (Figure 1.12).

1.8.7 Humdity

> ***Relative Humidity:*** *the vapour content of air (measured as vapour density or pressure) expressed as a percentage of the vapour content needed to saturate air at the same temperature* (McIlveen, 1986).

Relative humidity is a widely known measure of atmospheric moisture. It is analytically defined as "the ratio of the air's vapour pressure (e) to the saturated vapour pressure at the air's temperature e_s" (Linacre & Geerts, 1997), and is expressed as a percentage,

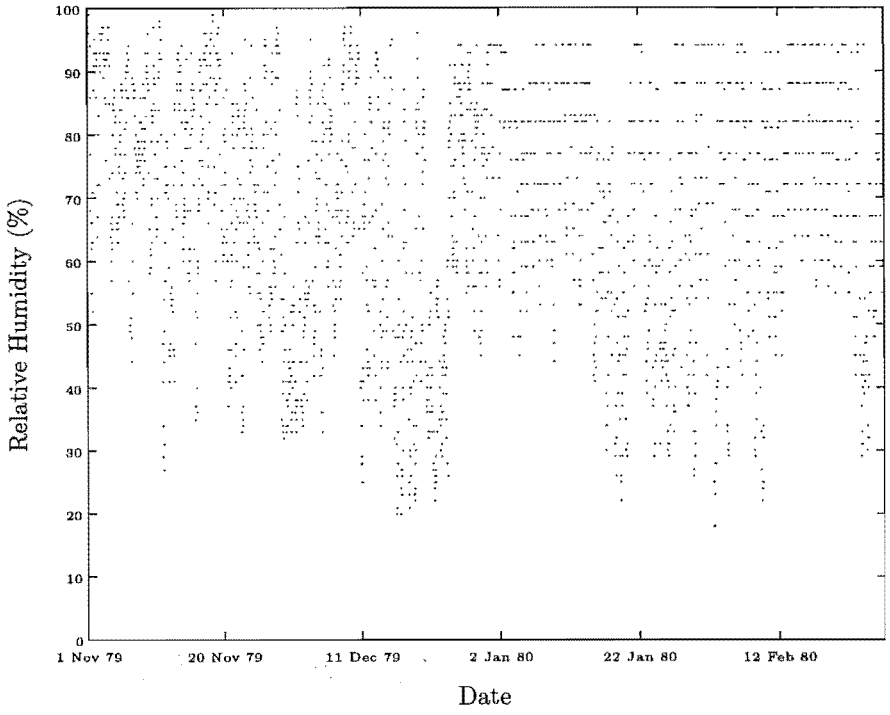


Figure 1.16: Hourly relative humidity values for November 1979 - February 1980.

$$RH = 100e/e_s\%. \tag{1.3}$$

Relative humidity is not measured directly, but is calculated from temperature and dewpoint measures. Hygrometric tables (New Zealand Meteorological Service, 1961) are used to calculate relative humidity from hourly spot readings of wet- and dry-bulb temperatures.

Hourly humidity measures were recorded to the nearest percent from 1968—1989, and again for 1998—1999, while values were given to one decimal place for the 1990—1997 time frame. There were no missing values in the relative humidity series as a result of the analytical nature of the variable. Due to NIWA using a two-digit coding system (00-99), values recorded as '00' represent 100% relative humidity and were therefore recoded as such. These 100% relative humidity values occur when the dewpoint equals the temperature.

At an hourly level, the relative humidity values had three obvious changes in the series, though the range of values remained consistent. Figure 1.16 illustrates the change in January 1980, when the recording method was updated from hygrometric tables to a computerized calculation. Up to 1980, the data were distributed regularly throughout the range of values, but from 1980—1990, there appeared to be a predominance of values recorded at certain levels, approximately 5% apart for those values greater than 60% relative humidity. From 1990—1993, this trend was not as obvious though still remained evident, and from mid-1994 the recordings returned to their initial distribution. The 1990 change directly reflects upgraded instrumentation used in the measurement of temperature and dewpoint, whilst it is speculated that the 1994 change may have been a result of upgraded computer software. These discrepancies are not evident in the daily summary variables.

Daily summary variables of minimum, mean, maximum, standard deviation, range and AHC measures were calculated for relative humidity (*Humid*) as detailed in Section 4.1 with regard to temperature. Gross seasonal fluctuations were evident in the minimum, mean, standard deviation and range series: relative humidity was higher in winter than in summer (on average, 80.9% and 70.7% for winter and summer

respectively). Maximum relative humidity obviously peaked at 100%, though there were some low values within the series. All three absolute hourly change variables appeared to oscillate within a regular range, with occasional spikes appearing randomly in the series (not shown).

Figure B.10 presents the cumulative centile distribution for mean daily humidity in both Period 1, and Period 2, categorised into days where no SIDS occurred and days with at least one SIDS death. In Period 1 (Figure B.10 (a)), the humidity appears greater on days with no SIDS deaths, up to a humidity level of approximately 70%, when the two distributions cross. From that point on, the humidity is greater on days where at least one SIDS death occurred. In Period 2, (Figure B.10 (b)), there appears to be little difference in the two groupings for lower levels of humidity, but again, from approximately 70% humidity, SIDS days appear to have slightly higher values of humidity than non-SIDS days. This suggests that there may be an underlying association between humidity and the incidence of SIDS, although seasonality was not accounted for in the depicted data.

1.8.8 Pressure

> ***Pressure:*** *the apparently continuous and isotropic force exerted on unit area of any real or imaginary surface because of bombardment by molecules of contiguous fluid* (McIlveen, 1986).

Due to humans unresponsiveness to inaudible naturally occurring pressure variations, it was not until the mid seventeenth century, and the development of the barometer, that the existence of considerable, but much slower, changes in atmospheric pressure were discovered.

Spot readings of atmospheric pressure were taken hourly by a precision aneroid barometer. In 1990, the international standard units for pressure changed from pascals to hectapascals (HPa). The NIWA pressure readings changed at this point from being measured in kilo-pascals (HPa$\times$10) to HPa, and so the 1968—1989 readings were scaled accordingly to give a constant unit of measurement over the study time frame.

Apart from this scale change, the pressure series is consistent over the study period. Figure B.11 shows hourly readings of pressure for the years 1988—1989. The pressure values appear to oscillate almost at random, though there is a vague underlying seasonal pattern under these fluctuations.

Daily summary variables were calculated for pressure (*Pres*) as detailed previously. Of the 280 512 hourly measures, 98 were missing (approximately 0.03%), and ignored. The vague seasonal periodicity seen in Figure B.11 was replicated in the minimum, mean and maximum series, while the remaining measures were almost random (not shown).

The cumulative centile distribution for mean daily pressure, split into SIDS and non-SIDS days, in both Period 1 and Period 2, is presented in Figure B.12. The distribution of pressure is essentially the same for both SIDS days and days where no SIDS deaths occurred, possibly indicating little association between SIDS and daily pressure in either period.

1.8.9 Rainfall

> ***Rainfall:*** *a measure of the total rainfall, including drizzle and melted hail and snow.* (McIlveen, 1986).

Three categories of rainfall intensity are defined: slight (0.01 to 1.1 mm/hr), moderate (1.2 to 5.9 mm/hr) and heavy (> 6.0 mm/hr). Rainfall was recorded as the total rainfall in millimetres per hour (mm/hr), and measured (to the nearest 0.1 mm) using two gauges through the study period: an auto-daily rain gauge until July 1994, and then a hydra-drop gauge for the remainder of the study period. The rainfall pattern was consistent over the study period and not affected by the gauge change. There was one period of extremely heavy rain in early in 1975 when there were consecutive recordings of 23 mm/hr and 26 mm/hr. The

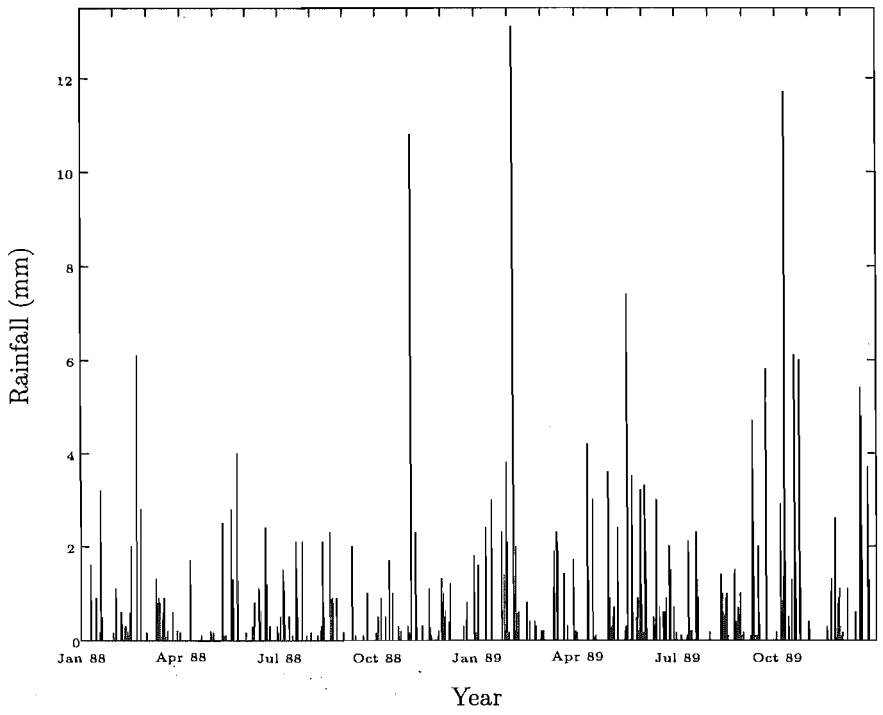


Figure 1.17: Hourly rainfall measures for 1988—1989.

majority of hours over the 32 years had no rainfall (91.4 %), as rainfall occurred in relatively short bursts. This pattern is highlighted in Figure 1.17, which presents the hourly rainfall readings for 1988 and 1989. There does not appear to be any seasonal distribution to the hourly rainfall data.

Daily summary variables were created for rainfall (*Rain*) as previously detailed in Section 1.8.2, with corresponding lags and between day effects. Minimum rainfall per day was replaced with a total rainfall per day measure, as with only 23 days (0.2%) where the minimum rainfall per day was not zero, the minimum daily rainfall provided little information. As with the hourly data, the daily rainfall summary variables showed little seasonal dependency (not given).

Figure B.13 presents the cumulative centile distribution for total daily rainfall in Period 1 and Period 2, again grouped into SIDS and non-SIDS days. From Figure B.13 (a), it appears that in Period 1, the total rainfall was higher on days with no SIDS deaths as compared to days where at least one SIDS occurred. The opposite trend is evident in Period 2 (Figure B.13 (b)). This highlights a potential relationship between daily rainfall and SIDS incidence, which will be investigated fully in the following chapters.

1.8.10 Sunshine

> ***Sunshine:*** *direct solar radiation, as distinct from radiation that has been diffused by clouds or the atmosphere* (Dunlop, 2001).

Seasonal affective disorder (SAD) is a pattern of depression that is related to changes in seasons. Depression occurs in the duller autumn and winter seasons but lifts with the onset of spring (Nevid et al., 2000). The cause of SAD is unknown, but one theory suggests it is a result of deficiencies in transmitting serotonin in some parts of the central nervous system during winter (Schwartz, 1997). As variability in sunshine has such an effect on humans, there is potential for an underlying relationship between sunshine and the incidence of SIDS, either directly, or indirectly through changes in parental behaviour.

Sunshine values were only available at a daily level for the period of the study. Recordings are the total sunshine hours per day, measured to the nearest tenth of an hour. The complete series is visually consistent (not shown) and there are no missing values. Figure B.14 illustrates the total sunshine hours per day for the years 1988—1989. A cyclic pattern is evident with the maximum sunlight hours peaking in summer and troughs corresponding to shorter days in winter. There appears little evidence of non-random patterns in the low values of sunlight.

As the sunshine data was not available on an hourly scale, only 42 day effects, between day effects and lags were created. These include the total number of sunshine hours per day for *day*0 and days up to fourteen days prior (*day*-14), the difference in total sunshine hours per day on consecutive days, and from *day*0 and previous lags, past average day effects and the difference between the total hours of sunshine on *day*0 and previous averages.

The cumulative centile distribution of the total sunshine hours per day in Period 1 is shown in Figure B.15 (a), where the data has been grouped into SIDS and non-SIDS days. The distribution for Period 2 is in Figure B.15 (b). In Period 1 there appears to be little difference in the sunshine hours per day for SIDS and non-SIDS days, with many crossovers between the two distributions. In contrast, in Period 2, there is a distinct difference in the distribution of the two groups, with there being less sunshine on days where at least one SIDS occurred. This indicates a possible dependency between SIDS and sunshine, although no seasonal effects were accounted for in Figure B.15.

1.8.11 Solar Radiation

> ***Solar Radiation:*** *electromagnetic radiation from the sun, especially in the wavelength range 0.3 to 3 μm which contains nearly all the total irradiance* (McIlveen, 1986).

The geometry of the Earth's motions determine the amount of radiation (or radiance) that the Earth receives from the sun. As the orbit of the Earth around the sun is elliptical rather than circular, the distance between the Earth and the sun varies from 147.1 million kilometres (the perihelion) to 152.1 million kilometres (the aphelion). 6.5% more radiation is received at the Earth's surface at the perihelion (in January) than at the aphelion (in July). With the temperature on the Earth largely determined by the amount of radiation it receives from the sun, the Earth's orbit tends to make global temperatures warmer in January than in July, that is, the Southern hemisphere summer should be hotter than summer in the Northern hemisphere. However, this trend is not observed due to the influencing effects of the different amounts of ocean in the two hemispheres (Linacre & Geerts, 1997).

The total global solar radiation in the hour leading up to the time of observation, (Mj/m^2) is a measure of light energy from the sun. Four different radiometers were used to measure this over the time frame of the study: up to 1974 an *Eppley 10-J Bulb-type-Global radiometer* was used, and then an *Eppley 8-48A BWP-Global radiometer* till 1978. From 1978—1984, an *Eppley PSP-Global radiometer* was used, and the final radiometer was a *Li-Cor Silicon Cell.*

As with sunshine, solar radiation values were only available at a daily level, with recordings being the total global solar radiation per day. There were no obvious changes in the pattern of the series over the time frame of the study. This daily radiation measure is illustrated in Figure B.16 for the 1988—9 years. From this figure, a strong seasonal trend is evident, with peaks in summer and troughs in winter (on average, solar radiation is 20.98 Mj/m^2 per day in summer and 5.80 Mj/m^2 per day in winter). The range of recordings appears larger in the summer season, than the winter.

Day effects, between day effects and lags were created for solar radiation (*Rad*) in the same way as for sunshine, giving 42 daily summary measures. There were 88 days with no recording (0.75%), that is, missing

Reading	Pattern
Less than 50m	Dense fog
50m but not 200m	Thick fog
200m but not 500m	Fog
500m but not 1000m (1km)	Moderate fog
1km but not 2km	Mist, haze or very poor visibility
2km but not 4km	Poor visibility
4km but not 10km	Moderate visibility
10km but not 20km	Good visibility
20km but not 50km	Very good visibility
50km or more	Excellent visibility

Table 1.5: Visibility Readings and corresponding interpretation.

data. These missing values were replaced with the average solar radiation value for the day of the year corresponding to the missing value (averaged over all the years with data available for that day).

Figure B.17 presents the cumulative centile distribution for the total solar radiation per day in both Period 1 and Period 2, categorised into days where no SIDS occurred and days with at least one SIDS death. As with sunshine, there is an obvious difference between the distributions of the SIDS and non-SIDS days in Period 2 (Figure B.17 (b)). The total solar radiation is lower on days with at least one SIDS, indicating a potential relationship between solar radiation and the incidence of SIDS. This trend can also be seen in information presented for Period 1 (Figure B.17 (a)), though not as strongly. These relationships will be examined further in the following chapters.

1.8.12 Visibility

> ***Visibility:*** *The greatest distance at which an object with specific characteristics is visible to the naked eye* (Dunlop, 2001).

Visibility is essentially a measure of the distance that the observer can visually see. It is affected by both rain and fog but can also be affected by mist, haze, and salt particles of oceanic origin, as well as by pollution.

Spot readings of horizontal visibility were taken hourly and can be interpreted using the key given in Table 1.5. Readings range from a *dense fog* where visibility is less than 50m (0.05km) to *excellent visibility* defined by visibility greater than or equal to 50km.

During daylight hours, visibility was measured by an observer sighting various markers at known distances around the measuring station. The sighted marker that was the furthest distance from the observer corresponded to the measurement of visibility for that hour. Ideally measuring visibility during darkness should involve this same procedure but with illuminated markers. In reality, this rarely happens and the nocturnal measurements at Christchurch International Airport were estimated by contacting the air traffic control tower and querying the controller as to how far down the runway the runway lights were visible. This value was then converted to a visibility reading[†]. Visibility is therefore a very subjective measure, largely dependent on the observer.

Visibility was examined for its potential use as a very crude pollutant indicator, as a concomitant measure of pollution was unavailable for the time frame of the study. This follows the work of Knöbel et al. (1995) who used visibility as an optimetrical measure of air pollution.

[†]Personal communication, Allan Ryan, March 2002.

In an attempt to validate the use of visibility as a proxy for pollution, visibility was correlated with a small time series of Christchurch pollution data (28 May 1988 to 1 September 1988, and 7 May 1989 to 15 August 1989). Information on SO_2, NO, NO_2, CO and PM10 was available. The resulting correlation coefficients were very small (in the range of $r = 0.01$ to $r = 0.07$) implying no linear relationship between visibility and pollution. Visibility was therefore considered unsuitable as a proxy for pollution, and was not used further in the analysis of the climatic dependencies of SIDS deaths.

1.8.13 Dewpoint

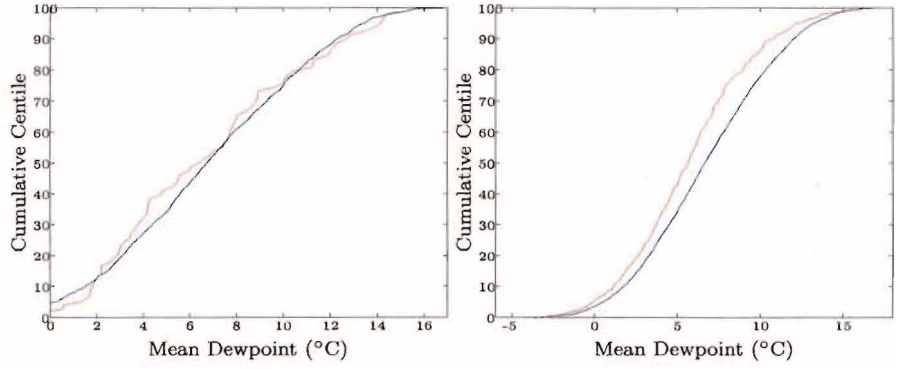
> ***Dewpoint Temperature:*** *the temperature at which a given parcel of air becomes saturated during isobaric cooling with conservation of vapour content* (McIlveen, 1986).

Another measure of atmospheric moisture content is dewpoint temperature, more commonly called dewpoint. Dewpoint is measured on a temperature scale and corresponds to "the temperature to which the air must be cooled for the moisture present to represent saturation" (Linacre & Geerts, 1997). Dew, or dewfall, is formed as water vapour from the net loss of long wave radiation, which causes water vapour in the atmosphere to condense onto chilled surfaces. This occurs when the saturated vapour pressure at the air's temperature (e_s) is less than the air's vapour pressure (e). The excess vapour, condensed into liquid, is then released as dew (Linacre & Geerts, 1997). As previously mentioned in Section 1.8.7, there is a close relationship between dewpoint and relative humidity, and conversion can be made between the two, and other measures of humidity, by using hygrometric tables (New Zealand Meteorological Service, 1961).

Dewpoint measures (in °C) were calculated from hourly spot readings of wet-bulb temperature measures, using hygrometric tables for conversion. Over the 32 year series, there were three obvious changes in the hourly dewpoint data, with the first two changes occurring at the same time points as those in the temperature series. This was due to the close relationship in measurement technique between temperature and dewpoint. The points of change correspond to a unit change from degrees Farienheight to degrees Celsius at 1970 while in 1990 and 1994, they correspond to upgraded instrumentation and digitalisation, providing a finer measurement scale. Figure B.18 presents the hourly dewpoint data for the years 1988—9. The underlying seasonal periodicities in the series appear blurred by random fluctuations about this annual cycle. As with temperature, the dewpoint series peaks in summer with readings of 10.15°C on average, as compared to 3.03°C in winter.

Daily dewpoint (*Dew*) summary measures were calculated for the same variables as detailed previously. Over the 32 years of data, there were 503 missing values. Values recorded as '50' correspond to hours where the wet-bulb was coated with ice, and were also considered missing. This occurred on 265 occasions. There was one obviously misrecorded value in the series of -38°C, occurring on 5 March, 1978. With the minimum dewpoint recorded as -13.7°C, this value was also considered missing. This gave a total of 0.27% missing data from the hourly series, which were ignored in the calculation of the daily summary values.

The cumulative centile distribution of mean daily dewpoint, categorised into days where no SIDS occurred and days with at least one SIDS death is presented in Figure 1.18. In Period 1 (Figure 1.18 (a)), there is some separation between SIDS and non-SIDS days, with dewpoint appearing lower on days where at least one SIDS had occurred for lower values of dewpoint. Figure 1.18 (b) shows clearly that the dewpoint was lower on days where at least one SIDS had occurred, as compared to days with no SIDS deaths in Period 2. This indicates a potential relationship between SIDS occurrence and dewpoint, although no seasonal effects have been accounted for in the information depicted in Figure 1.18.



(a) Period 1 (b) Period 2

Figure 1.18: Cumulative centile distribution for daily mean dewpoint (— SIDS days, — non-SIDS days).

1.8.14 Seasonality Indicators

> ***Seasonality**: the annual cyclic variation in the incident rate of SIDS. Typically this consists of a peak in the number of SIDS deaths in winter, with a corresponding low in summer.*

Seasonality has been accounted for in a variety of ways when examined in relation to the periodic distribution of SIDS. For example, Campbell (1989) used an indicator variable ranging from one to twelve to represent the months of the year; McGlashan & Grice (1983) used the monthy average temperature; similarly, Schluter et al. (1998) modelled the seasonal profile of SIDS using a retrospective moving average of monthly temperature; various sinusoidal functionals have also been utilised as seasonality measures (Douglas et al., 1998; Campbell et al., 2001).

Various daily seasonality measures are defined for the Canterbury climate data set, alongside variables that are independent of meteorological measures. From these, the variable which best describes the annual cyclicity of the daily Canterbury SIDS data will be determined (see Sections 3.4.1 and 3.4.2). This determination is a purely objective procedure, where the best measure of seasonality is chosen based upon the resulting goodness-of-fit statistics. The seasonality measures calculated are detailed below.

1. **Sinusoids**: $sin(\frac{2\pi t}{365}) + cos(\frac{2\pi t}{365}) \quad t = 15, \ldots, 11688$
2. **Sinusoids with second harmonics**:
   $sin(\frac{2\pi t}{365}) + cos(\frac{2\pi t}{365}) + sin(\frac{4\pi t}{365}) + cos(\frac{4\pi t}{365})$
3. **Month indicator variable**: $i = 1, \ldots, 12$
4. **Retrospective climatic average**: A retrospective 30-day moving average of a daily climatic variable, x, where the daily climatic variables to be considered are: minimum, mean and maximum daily measures of $Temp$, $WindS$, $North$, $East$, $South$, $West$, $Humid$, $Pres$, Dew and also total rainfall per day ($Rain$), total number of sunshine hours per day (Sun) and total global solar radiation per day (Rad). Averages are taken over the day of observation ($day0$) and the previous 30 days.
5. **Monthly climatic average**: Monthly average of daily climatic variable x where x corresponds to the daily climatic variables detailed previously in (4) above, but the average is now taken over month i (that is, not retrospectively).

1.9 Thesis Overview

Historically, statisticians and epidemiologists have played an important role in SIDS research. By examining information from large numbers of infants dying of SIDS, risk factors have been identified that occur more often in SIDS infants than in surviving infants. This epidemiological approach has resulted in modifiable risk factors being identified, which, through public health 'back-to-sleep' campaigns, directly resulted in a sharp drop in SIDS incidence in the early 1990s. However, this does not help determine whether an individual infant will die from SIDS; an infant may score highly with many known risk factors, and not succumb to SIDS, yet an infant who is considered 'low risk' may still die. The epidemiological approach offers a starting point for identifying aetiological factors essential in understanding the cause of SIDS. Historically, Canterbury has had one of the highest rates of SIDS in New Zealand, and New Zealand has had one of the highest SIDS rates in the western world (Nelson, 1996).

The aim of this thesis is to describe the incidence of SIDS in Canterbury, New Zealand (1968—1999), in terms of complex weather patterns, characterised by a diverse array of climatic variables. This is achieved by linking the temporal sequence of SIDS counts with a comprehensive climate profile. This study is the first to systematically analyse a multiplicity of climate data at different temporal levels (hourly, daily and monthly), with an accurate extensive time series of SIDS.

A unique chronological profile of SIDS counts in Canterbury, over the years 1968—1999, was analysed. This SIDS data is distinctive as it is unusual to have a series spanning such a long time period that does not suffer from changes in diagnostic policy. Autopsy and pathology records were examined retrospectively to accurately ascertain the cause of death (Ford, 1986), leading to a comprehensive, complete ascertainment of all SIDS deaths in the area over the time of the study.

The association between climate and SIDS has a long history, with the first reference to a seasonal variation in SIDS (a peak in the incidence of SIDS in colder months) published nearly 150 years ago (Wakley, 1855). Many studies have related SIDS to various meteorological measures throughout the world, yet the only consistent relationship found is between SIDS and seasonality (for example Douglas et al. (1998) or Mitchell et al. (1999)). Some studies have shown short term climatic relationships, over and above seasonal variations in the incidence of SIDS, including short term temperature measures (Campbell, 1994; Schluter et al., 1998), atmospheric pressure (Campbell et al., 2001), and pollution levels (Auliciems & Barnes, 1987; Knöbel et al., 1995). However these findings do not occur consistently across the SIDS literature.

This is the first comprehensive study of the incidence of SIDS with respect to climate, and the effects of differential lags, over such a wide range of climatic measures. Climatic variables considered in the analysis were derived from hourly measurements of temperature, wind direction, wind speed, wind velocity, wind chill, relative humidity, atmospheric pressure, rainfall, sunshine, solar radiation, visibility, and dewpoint, giving nearly 2.5 million weather observations. Comprehensive day, within day and between day measures and lags, were defined and examined in relation to the incidence of SIDS, over and above seasonality.

The analysis presented in this study aims to build a profile of high risk climatic patterns, and obtain results that will be useful in generating new hypotheses and future research directions. These will aid in the understanding of the underlying cause of SIDS, across regional, national and international levels.

This introduction provides an overview of the historical background and development of sudden infant death syndrome. The major epidemiological characteristics and risk factors for SIDS were discussed, alongside a detailed presentation of SIDS rates in New Zealand. The national SIDS profile was viewed with respect to changes in the neonatal and postnatal infant death rates over time, as well as highlighting ethnic and geographic trends. The Canterbury SIDS data set is described in this introduction chapter and is the primary focus of this study. Throughout the thesis the Canterbury SIDS deaths define the outcome variable in the various statistical analyses examined. The method of data collection is presented, alongside a description of the chronological profile of the Canterbury SIDS series. This chapter also presents an overview of the literature relating SIDS to various climatic patterns, alongside descriptions and details pertaining to the

creation of approximately 2500 climate variables, which are examined for potential relationships with the incidence of SIDS.

Chapter 2 examines the temporal sequence of SIDS counts with the aim of identifying underlying structural shifts in the SIDS series. To date, analysis of SIDS time series has not involved statistical tests for change points in the series, yet to successfully achieve the aim of this study, the characteristics of the chronological SIDS series need to be incorporated into the climate modelling process. This chapter presents three distinct methods to locate significant points of change in the temporal sequence of SIDS counts, namely a Survival method, a Block Bootstrap method, and a Mixture method. All three methods use a prescribed discriminating statistic to identify significant change points in the series, where the statistic is characterised by the underlying parameterisation of the respective method. Any significant change points identified will be incorporated into further statistical analyses involving the Canterbury SIDS series. This will be achieved by examining the resulting periods, defined by the change points, as distinct temporal series, and modelling them separately.

In Chapter 3, the daily SIDS counts are examined in relation to the climatic variables using a logistic regression analysis. The daily SIDS outcomes are viewed as a dichotomous variable, defined by days where no SIDS occurred and days where at least one SIDS occurred.

Principal component regression models are presented in Chapter 4. Climatic components are created, which are then included as covariates in logistic regression models. The components are uncorrelated, yet structured so they capture the information contained in the original climate variables. This method eliminates the problem of strong interrelationships which are inherent within varying measures of the weather.

Chapter 5 describes the conversion methods used, and variables created, to summarise the daily SIDS and climate data into a monthly time frame.

Relationships at a monthly level between climate and the incidence of SIDS in Canterbury are examined in Chapter 6. The profile of monthly SIDS counts is assumed to follow a Poisson distribution and hence this chapter involves modelling using Poisson regression techniques.

Chapter 7 examines the relationship between SIDS and climate at a monthly level using five different Poisson based mixture model formulations. The model formulations are Negative Binomial, Finite Mixture, Zero-inflated Poisson, Hurdle, and Extended Poisson Processes. They are all extensions of the Poisson distribution formulation, yet characterise the underlying distribution of SIDS counts in slightly different ways. These five mixture models are able to highlight differential climatic effects between months where SIDS occur and months where there are no SIDS deaths, while retaining the underlying Poisson formulation inherent in the monthly SIDS counts. Utilising models of this type may better capture the climatic dependency of the monthly incidence of SIDS, enabling an enhanced understanding of the relationship between climate and SIDS. Mixture modelling in this form is uncommon in the general time series framework.

Two distinct formulations which naturally incorporate the time series profile of SIDS counts are presented in Chapter 8. The first section assumes that potential serial correlation evident in the SIDS series has arisen from a latent process. This latent process introduces both overdispersion and autocorrelation into the observed count process, and conditional on the latent process, the observed series is independent. Two methods are used to estimate regression parameters in the presence of the latent process. In addition, the existence of such a process is formally tested. The second section in Chapter 8 analyses the relationship between monthly SIDS incidence and climate using generalised additive models. Generalised additive models are a popular tool for regression type studies of climatic time series and health outcomes, yet, to date, have not been utilised in the study of SIDS epidemiology. Generalised additive models are a nonparametric alternative to generalised linear models that permit nonlinear functions of the climatic covariates to be incorporated into the regression model.

Conclusions are discussed in Chapter 9, with comparisons made between the methods presented in each chapter. Future research directions are proposed, and considered at an international level.

Chapter 2

Identifying Change Points in the SIDS Series

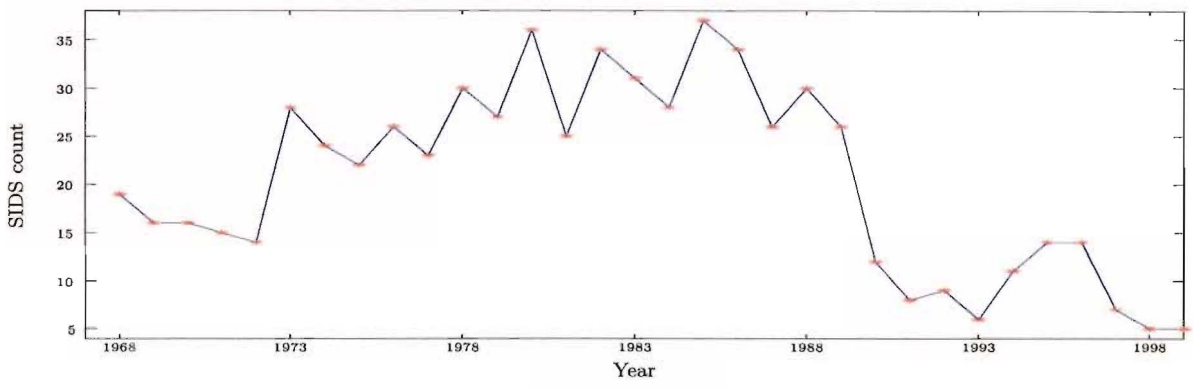
The overall aim of this study is to create a complete climatic profile of SIDS risk. To achieve this successfully, the characteristics of the chronological SIDS series need to be accounted for in the modelling process. The aim of the analysis presented in this chapter is to therefore examine the temporal sequence of SIDS counts in order to identify underlying structural shifts in the series. To date, analysis of SIDS time series has not involved statistical tests for change points in the series.

Figure 2.1 presents the chronological profile of SIDS counts at the three levels examined in this chapter: annually (Figure 2.1 (a)), seasonally (Figure 2.1 (b)) and monthly (Figure 2.1 (c)). The annual SIDS series was first presented in Chapter 1, Figure 1.9 (page 19), in relation to prone sleeping rates. At this level, the gross changes in the series are visually evident: the series appears to decrease slightly for the first five years of available data; there is a sharp increase from 1972 to 1973 (from 14 to 28 SIDS per year); SIDS numbers fluctuate around 30 per year from 1973 to 1988; there is then a steep decrease in the number of SIDS from 1988 to 1993 (from 30 to 6 SIDS per year); another small peak is evident in 1995—1996, with the SIDS numbers then dropping to two SIDS per year for the final two years of the study.

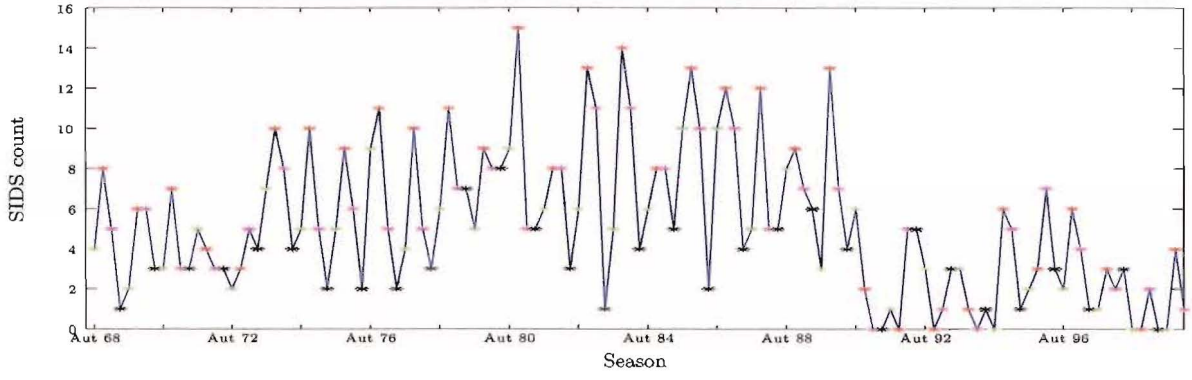
These trends are somewhat evident in the seasonal SIDS series, with the sharp decrease in SIDS numbers in the late 1980s visible. This seasonal plot (Figure 2.1 (b)) highlights the seasonal pattern underlying the series, with winter consistently recording the highest SIDS numbers, until the sharp decline in SIDS commencing at 1988. In the plot of monthly SIDS counts (Figure 2.1 (c)), these trends have been subsumed by the within year fluctuations of the series.

This chapter presents three distinct methods to locate significant points of change in the temporal sequence of SIDS counts. The methods, Survival, Block Bootstrap and Mixture, all identify significant change points by a prescribed discriminating statistic characterised on the underlying parameterisation of the respective method. No prior information regarding the location of potential change points is needed.

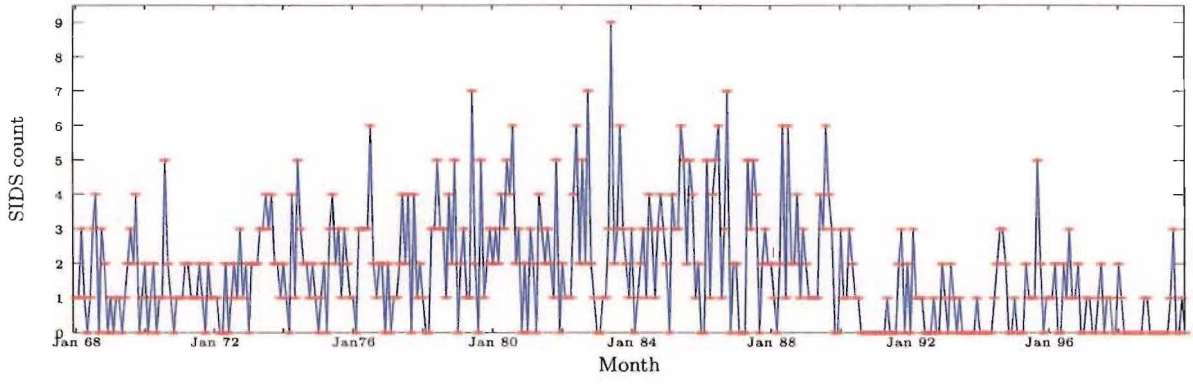
Three levels of temporal SIDS sequences are defined as follows:



(a) Annual SIDS counts



(b) Seasonal SIDS counts (⋆ summer, ◦ autumn, ∗ winter, · spring)



(c) Monthly SIDS counts

Figure 2.1: Time series of Canterbury SIDS counts, over different time scales, 1968–1999.

y_i^a = number of SIDS in year i
($i = 1, \ldots, 32$ corresponding to the years 1968—1999)
y_j^s = number of SIDS in season j
($j = 1, \ldots, 127$ corresponding to autumn 1968—spring 1999)
y_k^m = number of SIDS in month k
($k = 1, \ldots, 384$; from January 1968—December 1999).

The three methods presented each approach the change point problem in a unique way, with distinct underlying formulations. An iterative partitioning algorithm is utilised in both the Survival and Block Bootstrap methods, to locate all points of change in the SIDS series. Specific Matlab macros (*Version 6.1.0.1989a, Release 12.1; The MathWorks, Inc.*) were created for each method delineated below. This chapter follows the work presented by Dalrymple et al. (2001).

2.1 Survival Model

The first method utilised to examine the SIDS series for significant points of change uses an overdispersed survival model. This method is based on aggregate survival rates to a discrete point in time (Reed et al., 1998). Analysis is performed on the annual SIDS numbers and for the seasonal data, treating the four seasonal series separately. As seen in Figure 2.1, the gross changes visible in the series are swamped at both a monthly and seasonal level by the fluctuations within the year. The seasonal series is therefore split into four individual series, while this method of identifying change points was not considered suitable to be applied to the monthly SIDS series.

Let $\lambda^{(i)}$ be the hazard rate of SIDS in year i. Looking at the time to first SIDS occurrence, which clearly uses the conditional past profile (Kleinbaum, 1996), allows a SIDS cycle measure to be created. This measure can be interpreted as the expected time between SIDS, given that the current hazard rate prevails. The SIDS cycle is defined as the reciprocal of the hazard rate. Survival probabilities are then given by

$$q^{(i)} = \exp(-\lambda^{(i)}) \qquad i = 1, \ldots, 31. \tag{2.1}$$

If s_i = total number of SIDS after time i, the maximum likelihood estimate (MLE) for the survival probabilities in equation (2.1) is

$$\hat{q}^{(i)} = \frac{s_i}{s_{i-1}}. \tag{2.2}$$

The probability of a SIDS in the ith time period, conditional on no previous SIDS, is given by θ, where

$$\begin{aligned} \theta_i &= (1 - q^{(i)}) \prod_{a=1}^{i-1} q^{(a)} \qquad i = 1, \ldots, 31, \\ \theta_{32} &= \prod_{a=1}^{31} q^{(a)}. \end{aligned} \tag{2.3}$$

The null hypothesis (H_0), of no change points in the SIDS series, is tested against the hypothesis of a change point occurring at point p:

$$\begin{aligned} H_0: &\quad q^{(1)} = q^{(2)} = \cdots = q^{(n)} && (= q_0) \\ H_A: &\quad q^{(1)} = q^{(2)} = \cdots = q^{(p)} && (= q_1) \\ &\quad q^{(p+1)} = q^{(p+2)} = \cdots = q^{(n)} && (= q_2). \end{aligned}$$

The form of the alternative hypothesis (H_A) implies two distinct survival probabilities: q_1, the survival probability prior to p, and q_2, the survival probability after p. The test procedure uses a quasi-log likelihood function $Q(\cdot)$, derived from an overdispersed form of the multinomial distribution (McCullagh & Nelder, 1989). The corresponding likelihood ratio (LR) statistic is

$$R_p = \frac{2}{\hat{\sigma}_p^2}[Q(\hat{\Theta}_1) - Q(\hat{\Theta}_0)] \qquad p = 1, \ldots, 30, \tag{2.4}$$

where $\hat{\Theta}_1 = (\hat{q}_1, \hat{q}_2)$ (under H_A) and $\hat{\Theta}_0 = \hat{q}_0$ (under H_0), and $\hat{\sigma}_p^2$ is the Pearson estimate of the overdispersion parameter. Substituting the MLEs of q_1 and q_2 into equation (2.4) gives a LR statistic of the form

$$R_p = \frac{\nabla D}{\hat{\sigma}_p^2} = \frac{2}{\hat{\sigma}_p^2}\Big(\sum_{j=1}^{p}[s_j \log\frac{\hat{q}_1}{\hat{q}_0} + y_j \log\frac{1-\hat{q}_1}{1-\hat{q}_0}] + \sum_{j=p+1}^{n-1}[s_j \log\frac{\hat{q}_2}{\hat{q}_0} + y_j \log\frac{1-\hat{q}_2}{1-\hat{q}_0}]\Big). \tag{2.5}$$

The asymptotic distribution is χ_1^2, which is described below.

By viewing the SIDS series in terms of a counting process, the following variables can be defined:

$$\begin{aligned} N_1 &= \text{number of infants at risk pre change point } p \\ N_2 &= \text{number of infants at risk post change point } p \\ X_1 &= \text{number of infants surviving pre change point } p \\ X_2 &= \text{number of infants surviving post change point } p \end{aligned} \tag{2.6}$$

where N_j and X_j correspond to the risk set and event counts respectively. Then the associated survival probabilities, under the null hypothesis of no change point in the series, and the alternative hypothesis of a change point at p, can be written as:

$$\begin{aligned} \hat{q}_0 &= \frac{X_1 + X_2}{N_1 + N_2} \\ \hat{q}_1 &= \frac{X_1}{N_1} \\ \hat{q}_2 &= \frac{X_2}{N_2}. \end{aligned} \tag{2.7}$$

Therefore R_p from equation 2.5 can be rewritten as:

$$R_p = \frac{\nabla D}{\sigma^2} = \frac{1}{(N_1 + N_2)}\left\{\frac{\sqrt{N_2}(X_1 - N_1 q_0)}{\sqrt{N_1 \sigma^2 q_0(1-q_0)}} - \frac{\sqrt{N_1}(X_2 - N_2 q_0)}{\sqrt{N_2 \sigma^2 q_0(1-q_0)}}\right\}^2. \tag{2.8}$$

The event counts X_1 and X_2 can thus be viewed as overdispersed binomial random variables (McCullagh & Nelder, 1989), where, under the assumption of the null hypothesis, the mean and variance are given by:

$$\begin{aligned} E[X_j] &= N_j q_0 \\ Var[X_j] &= N_j \sigma^2 q_0(1-q_0), \end{aligned} \tag{2.9}$$

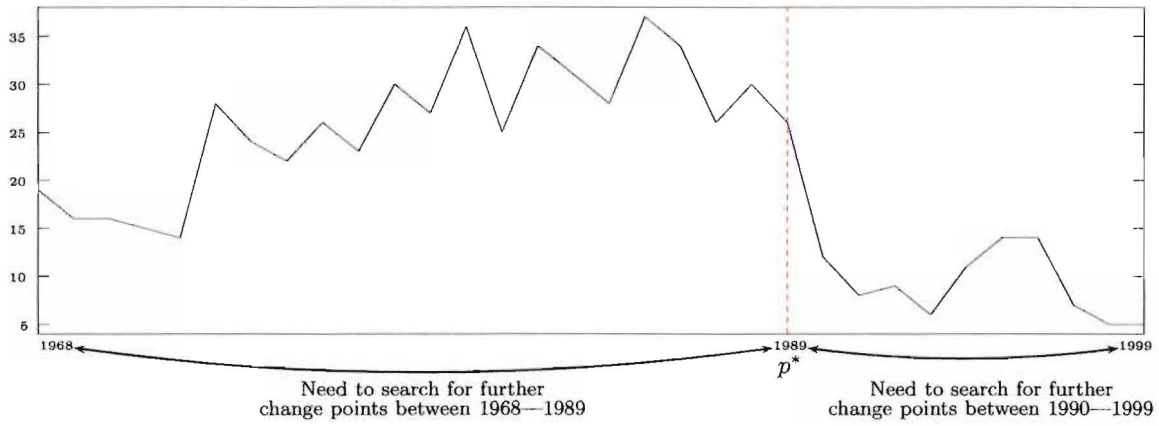


Figure 2.2: Iterative partitioning, to identify further change points in the series, when p^* is be significant.

where $j = 1, 2$, represent pre and post change point values. For large N_j, the two terms in the parenthesis in equation 2.8 above, are approximately normally distributed with zero mean and variances. Therefore, as $N_1, N_2 \to \infty$, the distribution of R_p tends to χ^2_1. In other words, the standard asymptotic null distribution of the likelihood ratio statistic for testing H_0 against H_A is

$$R_p \sim \chi^2_1.$$

The test statistic is then given by a scaled maximised version of the largest R_p (occurring at say, point p^*), which, for general n is

$$R^* = \frac{n-4}{n-3} \max_p \{R_p : p = 1, \ldots, n-2\} \sim \chi^2_1. \tag{2.10}$$

Approximate LR confidence intervals for $\hat{q}_1$ and $\hat{q}_2$ (given the significant change point at p^*), are obtained using numerical methods to solve the inequality

$$\{\hat{q}_1, \hat{q}_2 : R^* \leq \chi^2_{1,\alpha}\}, \tag{2.11}$$

with Reed et al. (1998) presenting details.

Identifying further change points is necessary if the estimated change point p^*, is found to be significant. An iterative partitioning algorithm is applied, which involves splitting the SIDS series into two partitions: pre-change $y_1, \ldots, y^*_p$, and post-change $y^*_{p+1}, \ldots, y_{32}$. Each partition is then re-examined for further change points, using the method outlined above. This procedure is illustrated in Figure 2.2, where a significant change point is shown at p^*, corresponding to 1989. The periods, 1968—1989 and 1990—1999 must then be examined further for any remaining significant points of change in the series. This partitioning continues until no further significant change points are located.

2.1.1 Change Points Identified via Survival Model

In the yearly SIDS time series, two significant points of change were found, 1979 and 1990. These effectively partition the annual series into three temporal blocks, each with a distinct survival probability, hazard rate, and hence different SIDS cycle, as summarised in Table 2.1. The middle block (1980—1990) has the lowest survival probability of 0.91 (95% CI: 0.88, 0.94), with an associated hazard rate of 0.10, and only 10 days

Block	$\hat{q}$ (95% CI)	Hazard rate(95% CI)	SIDS cycle (95% CI)
1968—1979	0.962 (0.954, 0.972)	0.039 (0.028, 0.047)	25.8 days (21.2, 35.2)
1980—1990	0.905 (0.883, 0.941)	0.100 (0.061, 0.124)	10.0 days (8.0, 16.5)
1991—1999	0.985 (0.970, 0.993)	0.016 (0.007, 0.030)	66.2 days (32.8, 142.4)

Table 2.1: Summary of survival type parameters for the annual SIDS series.

Annual	Summer	Autumn	Winter	Spring
1979	**1972**	**1972**	**1972**	**1972**
1990	1978	1975	1979	1981
	1982	1984	**1990**	1987
	1986	**1989**	1997	**1990**
	1990	1993		1997
	1993			

Table 2.2: Change points identified using the survival model for both annual and seasonal SIDS series.

expected between SIDS. This contrasts sharply with the final block (1991—1999), which has the highest survival probability of 0.99 (0.97, 0.99), with an associated hazard rate of 0.02, and an expected cycle of 66 days between SIDS.

Results for the change point analysis, by season, are summarised in Table 2.2. Up to six change points were located in one season (summer), which is possibly due to the influence of much smaller SIDS numbers when looking at seasonally grouped data. Only four change points were found in winter. There are two consistent years of change across all seasons: 1972 and 1989/90. This contrasts with the annual data analysis, where a change point was located at 1979 (also in winter) but not 1972. Identification of the change point at 1979, in the annual data, could therefore have been driven by the winter season where over 30% of the SIDS deaths occurred.

2.2 Block Bootstrap Model

Method 2 is a non-parametric change point analysis using a block bootstrap (Davison & Hinkley, 1997). This bootstrap method is applied to both the monthly SIDS series (y_k^m, $k = 1, \ldots, 384$) and the seasonal (temporally ordered) SIDS series (y_j^s, $j = 1, \ldots, 127$). This method is not used for the shorter time series of annual SIDS numbers, as grouping into years removes the seasonal component that Method 2 accommodates for by block bootstrapping.

Method 2 forms surrogate bootstrap series, under the null hypothesis of no change point, which give comparative data, with a similar autocovariance structure to the original series. R surrogate bootstrap series (y_k^{m*r}, y_j^{s*r}, r = 1,...,R) are created by randomly resampling the SIDS series in overlapping blocks. The block length (l) for the monthly data was chosen as $l = 30$, and for the seasonal data $l = 15$, so that the observed yearly cycle remains evident in the bootstrap data. After forming the surrogate series, any hypothesised change at p can be investigated by observing how closely the SIDS data match these surrogates.

The discriminating statistic, B_p (B_p^m for the monthly data and B_p^s for the seasonal data) is defined as the studentised difference in the mean SIDS rate, before and after p. The data set is truncated at each end to ensure that the comparison between data sections is based on reasonable sample sizes. For the monthly data the discriminating statistic is

$$B_p^m = \frac{(\bar{y}_1^{m,p} - \bar{y}_{p+1}^{m,n})}{\hat{\sigma}_n^m} \qquad p = 40, \ldots, n-40 \tag{2.12}$$

where $\bar{y}_a^{m,b} = (b-a+1)^{-1} \sum_{j=a}^{b} y_j^m$ and $\hat{\sigma}_n^m$ is the estimate of the standard deviation, using all 384 data points. Following similar notation, the discriminating statistic for the seasonal data is

$$B_p^s = \frac{(\bar{y}_1^{s,p} - \bar{y}_{p+1}^{s,n})}{\hat{\sigma}_n^s} \qquad p = 10, \ldots, n-10. \tag{2.13}$$

The bootstrap equivalent for equation (2.12) is

$$B_p^{m*} = \frac{(\bar{y}_1^{m,p*} - \bar{y}_{p+1}^{n,*})}{\hat{\sigma}_n^{m*}} \qquad p = 40, \ldots, n-40, \tag{2.14}$$

and similarly for equation (2.13). The bootstrap p-value for B_p^* is

$$pr^* = \frac{\#(B > B_p^*)}{R} \tag{2.15}$$

where # represents the number of times the specified event occurs. A small pr^*-value indicates a reduction in the mean SIDS rate after the change point, whilst a large pr^*-value indicates an increase in the mean SIDS rate after the change point, p. An iterative partitioning algorithm, analogous to that described in Section 2.1, is then used with the block bootstrap model, to locate all significant points of change in the SIDS series.

2.2.1 Cyclicity to Determine Change Points

An advantage of the bootstrap is its versatility in defining a different discriminating statistics, based on the sample autocorrelation function, which with block bootstrapping, allows investigation of changes in the periodicity (seasonality) of the monthly SIDS series. Bootstrap data is created under the null hypothesis of no change in cyclicity. The sample autocorrelation function at lag 12 is

$$r_{12}^{(a,b)} = \frac{1}{(b-a+1)^2} \frac{\sum_{k=a}^{b-12}(m_k - \bar{m})(m_{k+12} - \bar{m})}{\sum_{k=a}^{b}(m_k - \bar{m})^2}. \tag{2.16}$$

This allows a second discriminating statistic to be defined as the difference in the sample autocorrelation function

$$C_p^* = \hat{r}_{12}^{(1,p^*)} - \hat{r}_{12}^{(p^*+1,n)}. \tag{2.17}$$

2.2.2 Change Points Identified via the Block Bootstrap Model

The analysis of monthly SIDS data involved $R = 5000$ bootstrap replications. Figure 2.3(a) shows the studentised mean difference B_p^{m*} and Figure 2.3(b) the corresponding bootstrap p-value, pr^{m*} (the upper and lower significance limits of 0.025 and 0.975 are shown as dotted lines). The mean difference (and therefore pr^{m*}) first crosses the significance boundary at October 1985 and remains significant until September 1996. The maximum B_p^{m*} of 0.87 ($pr^{m*} = 0.00$) occurs at June 1990 ($p^* = m_{270}$), which is taken as the first change point in the series. The bootstrap confidence intervals are expected to be concave in shape as, when the change point p approaches 1 or n, one section of the data contains fewer results and is therefore more variable.

The monthly data was then partitioned, at p^* (June 1990), into two subgroups and each block examined for further change points using the block bootstrap technique. One further change point was identified at February 1973 ($p^* = m_{62}$, $B_p^{m*} = -0.63$, $pr^{m*} = 0.998$).

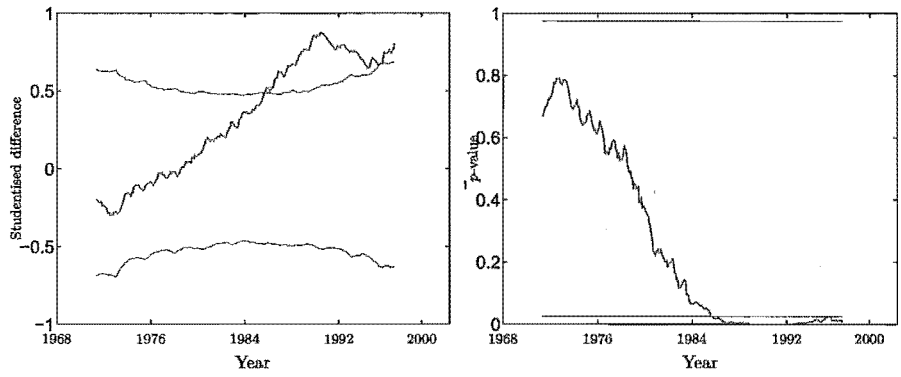


Figure 2.3: Observed mean difference and 95% CI for bootstrap difference and p-value for complete monthly series, 1968—1999.

As with the survival method, the monthly series appears to be made up of three distinct periods: January 1968—February 1973, March 1973—June 1990, and July 1990—December 1999. Figure 2.4 shows the monthly SIDS data, with these change points, and a smoothed estimate of the mean rate of SIDS using a boxcar kernel smoother (Diggle et al., 2002).

The analysis of the seasonal SIDS series located change points in similar temporal positions to that of the monthly series, namely at autumn 1973 and summer 1990.

Cyclicity to Determine Change Points

There is a strong yearly pattern in the monthly SIDS data, with more SIDS in winter and less in summer. This is particularly consistent in the middle period (March 1973—June 1990).

Results showed that the seasonal periodic cycles remained stable in the SIDS series until June 1990 ($p^* = m_{270}$). This confirms the first change point identified using the discriminating statistic B_p. The cyclicity results show a greater correlation in the data prior to p^*, as compared to that after, suggesting that the yearly cycle is not as strong after p^*.

2.3 Mixture Model

The final method of statistically identifying change points in the SIDS data involves viewing the series as a mixture of Poisson distributions. Mixture models have become increasingly popular in recent times, especially in application to disease mapping (Böhning, 1999; Böhning & Ayuthya, 1999; Böhning et al., 2000). Instead of viewing the data as a spatial map, this analysis looks at the SIDS series as a temporal map, applying a posterior Bayes classification model to determine points of change. Analysis is performed on the annual SIDS numbers and separately for each of the four seasonal series.

In the mixture model framework, autocorrelation in the data is assumed to arise from unobserved heterogeneity, where the population consists of k subpopulations due to an undiscovered explanatory variable. The term unobserved heterogeneity is used as initially, it is not known which subpopulation the observations come from (they are classified at the final step of analysis).

With the SIDS data, a Poisson density is assumed for each subpopulation j, with a potentially different parameter λ_j such that the density for subpopulation j is given by

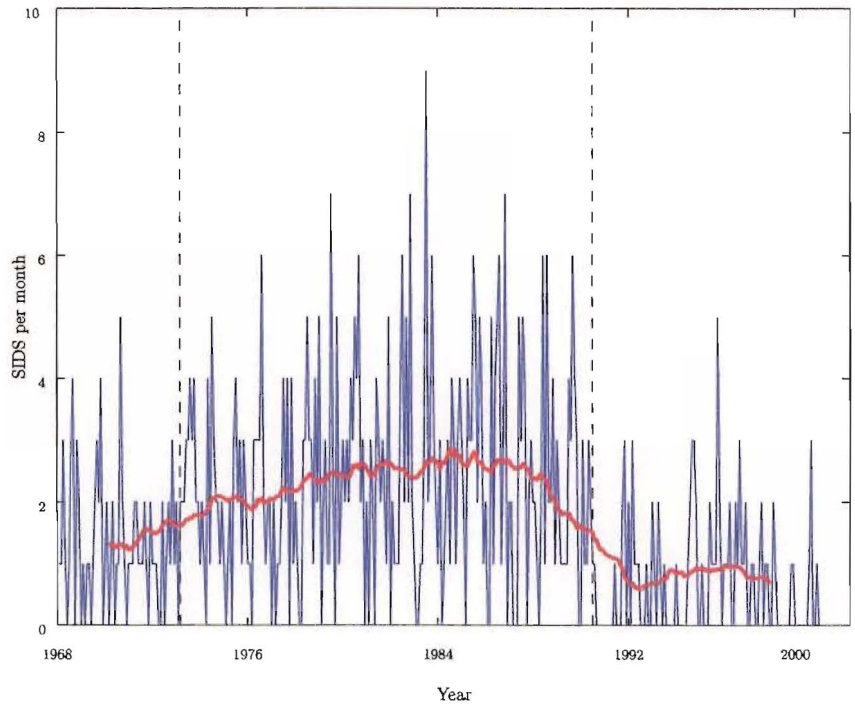


Figure 2.4: Number of SIDS per month with the smoothed estimate and significant change points for the block bootstrap method.

$$f(y, \lambda_j) = \frac{e^{-\lambda_j}(\lambda_j)^y}{y!} \qquad j = 1, \ldots, k, \qquad y = 0, 1, 2, \ldots. \tag{2.18}$$

Let $\mathcal{Z}$ be a latent variable describing population membership, the joint density $f(y, \mathcal{Z})$ is then

$$f(y, \mathcal{Z}) = f(y|\mathcal{Z})f(\mathcal{Z}) = f(y, \lambda_{\mathcal{Z}})pr_{\mathcal{Z}}, \tag{2.19}$$

where $f(y|\mathcal{Z}) = f(y, \lambda_{\mathcal{Z}})$ is the density conditional on membership in subpopulation $\mathcal{Z}$, $\lambda_{\mathcal{Z}}$ the parameter (mean) for subpopulation $\mathcal{Z}$, and $pr_{\mathcal{Z}}$ the probability of membership in subpopulation $\mathcal{Z}$. Conditional on $\mathcal{Z}$, the marginal density of equation (2.19) for a k-component mixture is

$$f(y, P) = \sum_{\mathcal{Z}=1}^{k} f(y|\mathcal{Z})f(\mathcal{Z}) = \sum_{j=1}^{k} f(y, \lambda_j)pr_j. \tag{2.20}$$

Equation 2.20 is a mixture distribution with kernel $f(y, \lambda)$ and mixing distribution

$$P = \begin{pmatrix} \lambda_1 & \ldots & \lambda_k \\ p_1 & \ldots & p_k \end{pmatrix}.$$

Estimation of P is achieved via maximum likelihood (Mood et al., 1974), where the log-likelihood is of the form:

$$L(P) = \sum_{i=1}^{n} \log f(y_i, P). \tag{2.21}$$

$\hat{P}$ is then defined as the nonparametric maximum likelihood estimator (NPMLE), which is estimated using the EM algorithm (Dempster et al., 1977).

Estimation of k is achieved by comparing goodness-of-fit statistics for various k. The three statistics utilised are

$$\begin{aligned} D &= -2L(\hat{P}) \\ AIC &= -2L(\hat{P}) + 2m \\ SC &= -2L(\hat{P}) + m\log(n), \end{aligned} \tag{2.22}$$

where m is the number of parameters in the model, n is the sample size, and $L(\hat{P})$ is the log-likelihood defined in equation 2.21. The k-component model which minimises these statistics is taken as the model which best fits the SIDS series. The three goodness-of-fit statistics (D, AIC and SC) are based on the log-likelihood function presented in equation 2.21, and are discussed in detail in Chapter 3 (Section 3.2, page 64).

Another approach for identifying k is a formal LR test, where the null hypothesis of a k-component model best fitting the data is tested against a $(k+1)$-component model. Titterington et al. (1985) showed that conventional results for the limiting distribution of the LR statistic (of the form in equation 2.21) do not hold for mixture distributions. Therefore this method is not applied to the problem of identifying k for the SIDS series.

A second approach for identifying the number of components underlying the SIDS series is used. This involves taking k as its maximum possible value, and then bootstrapping the data to calculate basic bootstrap confidence intervals (Davison & Hinkley, 1997) for $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$. Where the parameter confidence intervals for the $\hat{\lambda}$s overlap, the subcomponents are combined, thus reducing k.

	Number of components (k)			
	1	2	3	4
m	1	4	6	8
D	314.626	229.939	226.562	226.562
AIC	316.626	237.939	238.562	242.562
SC	318.092	243.802	247.356	254.288

Table 2.3: Goodness-of-fit statistics for various k-component models of the annual SIDS series.

2.3.1 Posterior Bayes Classification to Identify Change Points

After estimating k and the corresponding NPMLE $\hat{P}$, it is necessary to classify each observation into one of the k subcomponents of the mixing distribution. These subcomponents can be considered as disjoint classes into which the population is partitioned, and with the SIDS series, the classes can be interpreted as temporal partitions. Classification is achieved by means of the posterior distribution

$$f(\hat{\lambda}_j|y_i) = \frac{f(y_i|\hat{\lambda}_j)\hat{pr}_j}{f(y_i, \hat{P})} \qquad j = 1, \ldots, k; \quad i = 1, \ldots, 32 \tag{2.23}$$

such that y_i is classified into component a if

$$f(\hat{\lambda}_j|y_i) = \max_a f(\hat{\lambda}_a|y_i). \tag{2.24}$$

Once classification has been performed, change points in the SIDS time series can be identified as the break points between the natural groupings of observations.

2.3.2 Change Points Identified via Mixture Model

A two component model was found to best fit the annual SIDS series. Table 2.3 presents the three goodness-of-fit statistics for k ranging from one to four. The two component model ($k = 2$) rates best across both the AIC and SC statistics. The $k = 3$ and $k = 4$ models both returned a lower D value, but once these models were penalised for their complexity, the two component model was preferred. The log-likelihood for the four component model was identical to that for the three component model (as reflected in the values of D). This is a result of two of the four components in the model being essentially the same, reducing it to a similar form as the $k = 3$ model.

Table 2.4 presents parameter estimates, and corresponding bootstrapped confidence intervals for $\hat{\lambda}_i$, for both a two and a three component model. The confidence intervals, in the $k = 3$ model, corresponding to $\hat{\lambda} = 7.89$ and $\hat{\lambda} = 12.96$ overlap and thus the components were combined. The confidence intervals corresponding to the two component model show no overlap, therefore also indicating that $k = 2$ for the annual SIDS series.

The corresponding NPMLE for the two component model is

$$\hat{P} = \begin{pmatrix} 10.43 & 28.51 \\ 0.48 & 0.52 \end{pmatrix}. \tag{2.25}$$

$\hat{P}$, above, shows that the two components occur with approximately equal weighting, the first having a mean SIDS per year estimate of $\hat{\lambda}_1 = 10.43$ (95% CI: 8.08, 13.86) whilst in the second component the mean SIDS per year is nearly three times as much with $\hat{\lambda}_2 = 28.51$ (26.26, 30.85). Figure 2.5 shows the results of the

	$\hat{\lambda}$	$\hat{p}$	(95% CI for $\hat{\lambda}$)
$k = 3$	7.89	0.13	(5.71—12.18)
	12.96	0.35	(10.44—15.54)
	28.51	0.52	(26.26—30.85)
$k = 2$	10.43	0.48	(8.08—13.86)
	28.51	0.52	(26.26—30.85)

Table 2.4: Parameter estimates and corresponding bootstrap confidence intervals for $k = 2$ and $k = 3$ component models for the annual SIDS series.

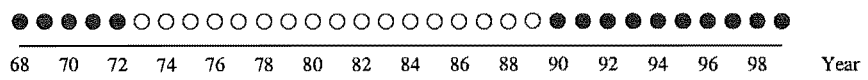


Figure 2.5: Posterior Bayes classification of the annual SIDS data. (• $\hat{\lambda} = 10.43$; ○ $\hat{\lambda} = 28.51$).

posterior Bayes classification, with the observations belonging to the first component represented by a solid circle. Two change points are clearly shown at 1972 and 1989.

Analysis for both the summer and autumn series found that one component was sufficient to fit the data (implying homogeneity) (summer: $\hat{\lambda} = 1.21$ (1.02, 1.40) autumn: $\hat{\lambda} = 1.50$ (1.33, 1.66)), whilst winter and spring were best modelled with two components (winter: $\hat{\lambda}_1 = 2.47$ (1.04, 3.92), $\hat{pr}_1 = 0.17$, $\hat{\lambda}_2 = 8.47$ (6.20,10.37), $\hat{pr}_2 = 0.83$; spring: $\hat{\lambda}_1 = 1.47$ (0.97, 4.23), $\hat{pr}_1 = 0.05$, $\hat{\lambda}_2 = 5.65$ (4.48, 6.94), $\hat{pr}_2 = 0.95$). The posterior Bayes classification for the winter temporal series highlighted identical change points to that found in the annual series (1972 and 1989), but only the later change point was identified in the spring data.

2.4 Change Points in the SIDS per 1000 Live Births Series

The SIDS rate series, SIDS per 1000 live births (SIDS/1000LB) was also examined for significant change points. The number of infants born did not remain stable over the study period (see Chapter 5, and Figure 5.2 for a discussion on the profile of births in Canterbury over this timeframe), and therefore may affect the location (or significance) of the change points found in the SIDS series. Figure 2.6 presents the SIDS/1000LB series, along with the pure SIDS counts, at an annual level for the years 1968—1999. The series have similar profiles, with the peaks and dips that occur in the SIDS series, reflected in the SIDS/1000LB series.

The three change point methods presented in the previous sections were rerun using the series SIDS per 1000 live births. At an annual level, the change points identified in the SIDS/1000LB series were identical to those previously noted in the SIDS series, that is 1972 and 1989. At a monthly level, change points were identified at very similar points in time for the SIDS/1000LB series to those previously found. A comparison of monthly change points for the SIDS and SIDS/1000LB series is given in Table 2.5 below.

The series SIDS/NAR was also tested at a monthly level, where NAR represents the number of infants at risk of SIDS in any one month (see Chapter 5 for full details of this variable), via the block bootstrap change point method. Change points were identified in the same months to those found in Table 2.5, for the SIDS/1000LB series.

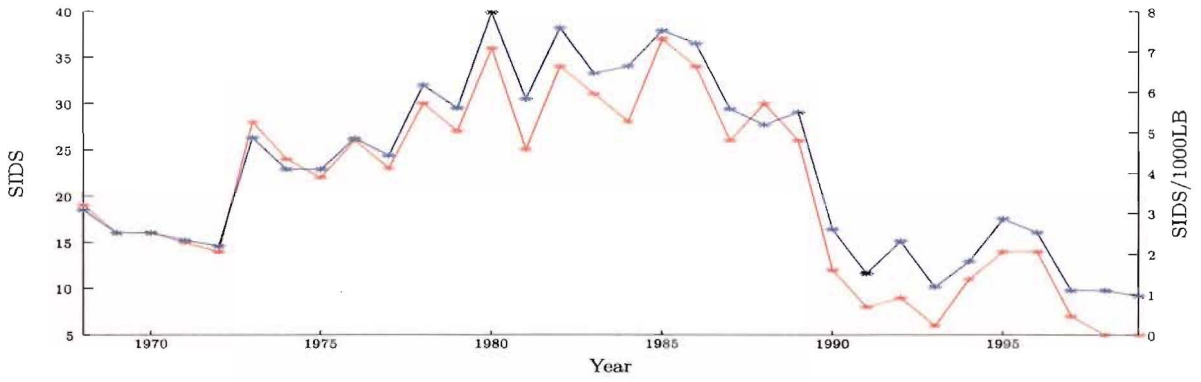


Figure 2.6: Annual SIDS profile with SIDS/1000LB, Canterbury 1968—1999 (* SIDS, * SIDS/1000LB)

Series	Change points	
SIDS per month	June 1990	February 1973
SIDS/1000LB per month	**May 1990**	**December 1972**
(Block-bootstrap test stat.)	$B_p^* = 3.46$	$B_p^* = -3.41$
	$(pr^* = 0.999)$	$(pr^* = 0.004)$

Table 2.5: Change points identified in monthly series (with corresponding test statistics and p-values for the SIDS/1000LB series).

2.5 Description of Three Temporal Periods

The three methods utilised to locate change points in the SIDS series returned similar results: two significant change points were consistently identified, located at 1972 and 1989. This effectively partitions the chronological profile of Canterbury SIDS counts into three distinct periods, namely 1968—1972, 1973—1989, and 1990—1999. Table 2.6, and Figure 2.7, present the mean number of SIDS per period, at both an annual level, and by seasons.

Table 2.6 shows that Period 2 has the highest mean number of SIDS per year. At 28.6 per year, this is nearly double the mean number of SIDS per year in Period 1 (16.0 per year) and more than triple the mean number of SIDS per year in Period 3 (9.1 per year). These trends are shown graphically in Figure 2.7. Interestingly, the number of SIDS in Period 3 is much more variable than those in Period 1, even though more SIDS were recorded in Period 1. This is reflected in the standard deviations of the number of SIDS per year which were 1.9 and 3.5 for Period 1 and Period 3 respectively.

Figure 2.7 shows that at a seasonal level, summer SIDS numbers remain reasonably stable across the three periods, with only a small increase in the mean number of SIDS in Period 2. In contrast, winter SIDS numbers follow the trends seen in the annual series. Here, Period 2 has the highest number of SIDS per year, double those in Period 1, and more than four times the mean number of SIDS in Period 3.

Figure 2.7 also clearly highlights the season pattern evident in the SIDS series, where on average, winter has the highest number of SIDS, and summer the lowest. Spring and autumn record similar SIDS rates, though spring records slightly more deaths than autumn in Period 1 and Period 2. This seasonal pattern is prominent in both Period 1 and Period 2 in Figure 2.7, but the mean number of SIDS per year, per season, in Period 3 is very similar; the highest rate (2.7 SIDS per year) was recorded in spring, and the lowest rate (1.8 SIDS per year) was recorded in autumn.

	Period 1 1968—1972	Period 2 1973—1989	Period 3 1990—1999
Annual	16.000	28.647	9.100
	(1.871)	(4.513)	(3.479)
Summer	2.500	3.900	2.100
	(1.000)	(1.919)	(1.729)
Autumn	3.200	6.412	1.800
	(1.304)	(2.093)	(1.874)
Winter	5.600	11.000	2.500
	(2.074)	(2.121)	(2.321)
Spring	4.400	7.412	2.700
	(1.342)	(2.123)	(2.406)

Table 2.6: Mean numbers of SIDS per period (standard deviations), at both annual and seasonal levels.

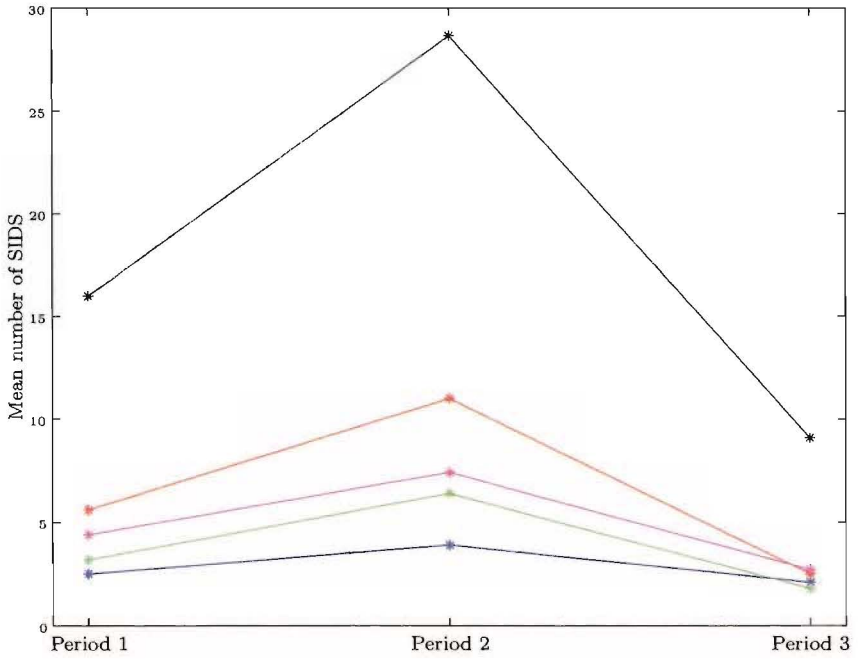


Figure 2.7: Mean number of SIDS per period, at both annual and seasonal levels (∗ annual, ∗ summer, ∗ autumn, ∗ winter, ∗ spring).

	Period 1		Period 2		Period 3	
No. SIDS	No. Months	% Months	No. Months	% Months	No. Months	% Months
0	13	22%	30	15%	63	53%
1	26	43%	39	19 %	34	28%
2	13	22%	49	24%	14	12%
3	5	8%	39	19%	8	7%
4	2	3%	21	10%		
5	1	2%	13	6%	1	1%
6			9	4%		
7			3	1%		
8						
9			1	0.5%		

Table 2.7: Frequency of the number of SIDS deaths per month.

2.6 Analysis involving Period 3

As discussed in the previous section (and Chapter 1), the incidence of SIDS has dropped dramatically in Period 3, post local 'back-to-sleep' campaigns. The mean number of SIDS per year in Period 3 is less than a third of that in the previous period (9.1 SIDS per year compared to 28.6 SIDS per year in Period 2).

Table 2.7 shows the frequency of SIDS deaths at a monthly level. In Period 3 53% of months recorded no SIDS deaths compared with 22% and 15% in Period 1 and Period 2 respectively. Yet, with a mean of 0.76 SIDS per month in Period 3, 53% of zero occurrences is not considered overdispersed.

Figure 2.8 presents the average number of SIDS per month for each of the three periods. Both Period 1 and Period 2 show a strong seasonal pattern, with deaths peaking in June – August. In contrast, the number of SIDS per month in Period 3 remains reasonably constant, with no evidence of the classic winter peak. This lack of seasonality in the Period 3 SIDS profile was confirmed by the cyclicity block-bootstrap analysis (see Section 2.2.1). This analysis showed stable seasonal cycles up to the change point separating Period 2 and Period 3. Poisson regression modelling of Period 3 failed to identify any significant relationships between the monthly SIDS series and seasonality variables (see Chapter 7 for details of Poisson regression modelling, and Section 3.14 for definitions of candidate seasonality variables). Exploratory analysis failed to find any relationship between the incidence of SIDS in Period 3 and climatic variables (as defined in Chapters 3 and 6).

With evidence of such sparcity of SIDS occurrences in Period 3 (just 91 SIDS over 3652 days), and a complete lack of seasonality in the series, the ability to identify statistically significant relationships in the Period 3 SIDS profile must be questioned.

Combined with the lack of occurrences of the event of interest is the lack of variability in the Period 3 series; more than half of the months where SIDS occurred only recorded one death. Given the low incidence rate, low variability, lack of seasonality and any evidence of climate relationships, Period 3 (1990—1999) will no longer be considered in the further analysis relating climate to the incidence of SIDS, as presented in this thesis. The Period 3 series is essentially a 'flat-line'; significant associations with a series lacking any profile cannot be identified with the statistical methods utilised in this thesis.

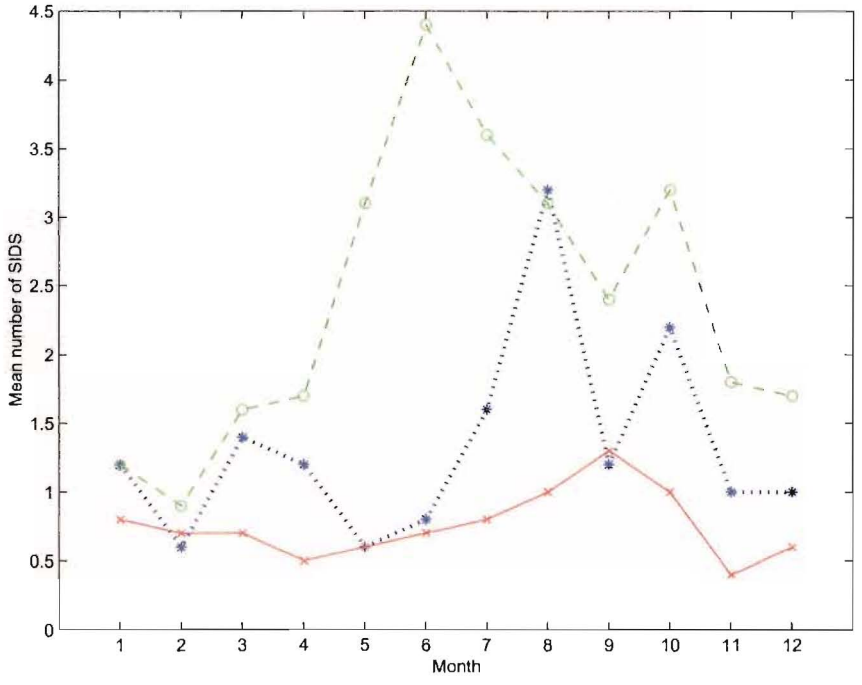


Figure 2.8: Mean monthly profile of SIDS by period (· · ∗ · · ∗ Period 1, 1968—1972; - - o - - o Period 2, 1973—1989; — x — x Period 3, 1990—1999).

2.7 Analysis of Period 1 and Period 2 data

Given the significant change point in the SIDS series at 1972, this section examines the question of whether to model the series by

(i) combining Period 1 and Period 2, and incorporating a dummy variable to model the period effect; or

(ii) model the Period 1 and Period 2 data separately as two distinct time series.

2.7.1 Profiles of SIDS in Period 1 and Period 2

Figure 2.9 presents the annual profile of SIDS counts for Period 1 and Period 2. As discussed in Section 2.5, the mean number of SIDS per year in Period 2 is nearly double that of Period 1 (28.6 SIDS/year compared with 16.0 SIDS/year). This level-shift is highlighted in Figure 2.9. The level-shift is confirmed by the change point analyses† which found a significant change point at 1972 (highlighted on Figure 2.9).

The level-shift is not the only difference between the profile of SIDS in Period 1 as compared to Period 2; the underlying cyclic variation in the incidence of SIDS also varies between the periods. This is graphically illustrated in Figure 2.8, which shows the average monthly profile for Period 1 and Period 2. In Period 1, the highest number of SIDS occurred in late winter to early spring. In contrast in Period 2, it was late autumn to early winter that recorded the highest number of SIDS deaths. This difference is essentially a three month 'shift' in the risk profile of SIDS, between the two periods.

According to bioclimatic models involving the Southern Oscillation index (ENSO) (Wolter & Timlin, 1998), Period 2 has been identified as a generally warmer and dryer period than Period 1 (A. Zeevi and R.

†The change point analyses are all based, in some form, on the number of SIDS before and after the change point.

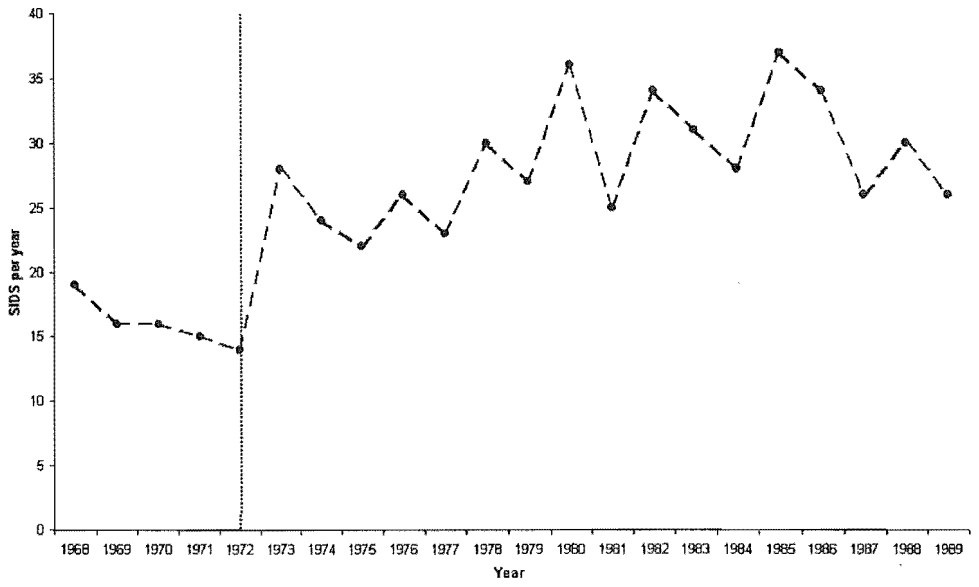


Figure 2.9: Profile of the annual number of SIDS in Canterbury, 1968–1989.

Meir and R. Adler, 1998). The observed three month shift in the risk profiles of SIDS (Figure 2.8) could possibly be explained by these different "climates" across periods. For example, the early peak in SIDS risk occurring earlier in the year in Period 2 could be a result of the early accumulation of heat, leading to hyperthermia.

2.7.2 Modelling Options

Given the level and seasonal shifts in the profile of SIDS in Period 1 and Period 2, three modelling options are considered as follows:

Model 1: $\text{SIDS} \sim \text{season} + \text{climate} + [\text{correlated}]\ \text{error}$

Model 2: $\text{SIDS}_1 \sim \text{season}_1 + \text{climate}_1 + [\text{correlated}]\ \text{error}_1$
$\text{SIDS}_2 \sim \text{season}_2 + \text{climate}_2 + [\text{correlated}]\ \text{error}_2$

Model 3: $\text{SIDS} \sim \text{season} + \text{climate} + \text{period} + \text{season}{*}\text{period} + \text{climate}{*}\text{period} + [\text{correlated}]\ \text{error} + [\text{correlated}]\ \text{error}{*}\text{period}.$

Models 1 and 3 are extremes, while Model 2 is a data driven compromise between them. Model 1 pools the data across time periods, assuming a constant seasonal profile. The presence of the level and seasonal shift in the SIDS data indicates that quite different profiles exsist between the two periods, casting serious doubt on the validity of Model 1.

Model 2 fits independent models to each of the periods. Model 3 is analogous to Model 2 above, and allows for heterogenous error structures across the periods, whilst modelling the two period profiles together.

If Model 3 cannot be reduced in any form, it can be assumed that no information can be shared by the two periods of data. Hence, fitting independent models of the form given in Model 2 would be appropriate.

It is assumed that 'Period' is a random effect. In the epidemiology literature, random effect models for incidence of air pollution with hospital admissions, where hospitals are viewed as random effects (REF), is typical of meta-analysis studies where treatments are compared across different centers (Brown & Prescott, 1999). The periods are viewed in an analogous way. Generalised linear mixed models, which incorporate random effects, are therefore used to fit Model 3, and for comparison.

2.7.3 Examining Modelling Options with Mixed Models

A generalised linear mixed model (GLMM) is fitted to the monthly SIDS data to examine Model 3 against Model 2 above. GLMMs extend generalised linear models (see Section 6.1.1, page 113) to include random effects, random coefficients, and covariance patterns (McCulloch & Searle, 2001). It is assumed that the monthly SIDS counts follow a Poisson distribution (see Chapter 6 for further details).

The GLMM fits a model of the general form below. Let $y_{i,j}$ denote the number of SIDS in the ith month and the jth period ($j = 1, 2$, $i = 1, \ldots, n_j$, $n_1 = 60$, $n_2 = 204$). Then

$$\begin{aligned} y_{i,j} &\sim \text{Poisson}(\mu_{i,j}) \\ \log(\mu_{i,j}) &= \mathbf{x}_{i,j}\boldsymbol{\beta} + u_j \\ u_j &\sim \text{iid } \mathcal{N}(0, \sigma_u^2), \end{aligned} \tag{2.26}$$

where $\boldsymbol{\mu}_j$ is the period specific mean, $\mathbf{X}_j$ the matrix of covariates, $\boldsymbol{\beta}$ the vector of regression coefficients and u_j the period specific random effects. This model uses a log link and normal distribution for the random period effects Brown & Prescott (1999).

The mixed effect model was implemented using the GLIMMIX macro in SAS (*SAS (R) Proprietary Software Release (8.01.01)*). The model was fitted using iteratively reweighted likelihoods (McCulloch & Searle, 2001). Random intercepts and period effects (modelling a level shift in the incidence of SIDS) were allowed for, alongside period-specific autoregressive (AR) error structures. Interactions between the period effect and seasonality measures and with climate variables were also examined. This permitted differential seasonality structures and climatic factors between Period 1 and 2.

Table 2.8 presents one such mixed model (chosen for illustration purposes), containing north-west wind velocity and dewpoint based climate variables, alongside seasonality and NAR (see Chapter 3 and Chapter 6 for full details on variable derivations and definitions). This model contains a significant period effect ($p = 0.001$), confirming the base level shift in the incidence of SIDS illustrated in Figure 2.9. A significant season by period interaction ($p = 0.02$), was also present, highlighting that differing seasonal components are needed to successfully model the pooled data. This confirms the significant difference in the underlying seasonal profiles across periods shown in Figure 2.8. Varying climatic profiles between the two periods are also in evidence in the mixed model presented in Table 2.8, with significant period by climate interactions ($p < 0.001$).

Thus Model 3, with the specific form presented in Table 2.8, cannot be reduced. Indeed the covariances and mean structures of the data vary across the periods; different time series (heteroskedastic AR errors), and different climate and seasonal components are required in the model. It is therefore correct to fit separate models, of the form given in Model 2, to each period[†]. In fact, the longer time profile in Period 2 (seventeen years compared to five years in Period 1), combined with the much higher incidence rate in Period 2, would result in Period 2 dominating any combined analysis.

[†]Personal communication, Dr James O'Malley, Harvard Medical School.

Variable		β (SE)
Fixed Effects		
NAR		0.04 (0.02)
Season	$\sin(\frac{2\pi t}{12})$	−0.32 (0.07)
	$\cos(\frac{2\pi t}{12})$	−0.49 (0.07)
Period*Season (1)	$\sin(\frac{2\pi t}{12})$	−0.01 (0.17)
	$\cos(\frac{2\pi t}{12})$	0.39 (0.17)
Climate	NW	1.43 (0.71)
	$Dew_{max(max)}$	0.01 (0.01)
Random Effects		
Intercept		0.02 (0.005)
Period (1)		−0.02 (0.005)
Period (2)		0.02 (0.005)
Period*Climate	NW (1)	−0.02 (0.005)
	NW (2)	0.02 (0.005)
	$Dew_{max(max)}$ (1)	−0.02 (0.005)
	$Dew_{max(max)}$ (2)	0.02 (0.005)

Table 2.8: Mixed model results for the years 1968–1989.

	Survival model SIDS cycle (days between SIDS) (95% CI)	**Block bootstrap model** Mean SIDS rate/month (95% CI)	**Mixture model** Mean SIDS rate/year (95% CI)
Period 1 (1968—1972)	40.8 (24.5, 82.8)	1.3 (1.2, 1.5)	10.4 (8.1, 13.9)
Period 2 (1973—1989)	11.6 (8.5, 15.4)	2.4 (2.2, 2.6)	28.5 (26.3, 30.9)
Period 3 (1990—1999)	66.2 (32.8, 142.4)	0.8 (0.6, 1.1)	10.4 (8.1, 13.9)

Table 2.9: Method and period specific summary statistics.

2.8 Summary and Conclusion

Three distinct methods of looking at the change point problem gave analogous results when applied to 1968—1999 SIDS data. Change points were located at 1972 and 1989, effectively partitioning the series into three periods. All analysis in this thesis will therefore incorporate these periods.

Table 2.9 summarises, by period, the main statistics found for each method. All parameters show the same general trend for SIDS numbers, namely that the middle period has at least twice as many SIDS occurring as the other two periods. The Survival method and Block Bootstrap method both show a slight decrease in SIDS numbers for the most recent period as compared with the first period.

The Survival analytic method, as a time to event analysis, accounted for the conditional time aspect of the series. Change points were defined as points of significant change in terms of hazards or SIDS cycles. The Block Bootstrap method used different definitions of the discriminating statistic to locate change points, and resulted in both the mean SIDS per month, and an estimated sample autocorrelation,

before and after the change points. The Mixture method of identifying change points assumed an underlying heterogeneity in the SIDS series. The distinct groupings were translated into a partitioned temporal map via a posterior Bayes classification. Of the three change point methods utilised, the Block Bootstrap method was the only method that accommodated the complete monthly SIDS time series, in that block sampling replicated the annual fluctuations in the series. Both the Survival and Mixture methods became insensitive to locating significant change points in the presence of within year seasonal fluctuations. The Mixture analysis specifically accounted for the discrete nature of SIDS with its assumption of Poisson counts.

An advantage of all three methods is that no change point needs to be pre-specified or known a priori. In essence, the methods scan over the complete time series looking for possible points of change. It is of interest to note that the change point at 1989 coincides with the back-to-sleep campaigns discussed in Section 1.7 (page 18). Visually, a sharp drop in SIDS numbers corresponding to this point has already been noted in Figure 1.9 (page 19) and Figure 2.1. The change point methods statistically confirm this visual observation.

Given the low incidence rate, low variability, lack of seasonality and any evidence of climate relationships, no further analysis will be performed with respect to relating climate to the incidence of SIDS in Period 3. Identification of statistically significant seasonal trends, and other climatic patterns cannot be successfully achieved when the SIDS death series lacks any profile.

Chapter 3

Logistic Regression Analysis of Daily SIDS Counts

Summary climate variables, ranging from temperature to wind chill, were introduced, defined and described in Chapter 1.8. This chapter examines these measures, at a daily level, in relation to SIDS occurrences in Canterbury. This is a preliminary analysis, with the overall aim of reducing the many potential covariates to a subset of variables, excluding those with no obvious relationship to SIDS. A secondary aim of this analysis is to begin to obtain an understanding of the complex relationship between SIDS and climate. This is achieved by examining trends, patterns and influences within those climate variables identified as being related to the incidence of SIDS. This initial analysis is completed using logistic regression on SIDS and climate data recorded at a daily level, with each period analysed separately (Period 1, 1968—1972 and Period 2, 1973—1989). Period 3 was not included in this analysis, as it was shown in Chapter 2, Section 2.6 that SIDS deaths in Period 3 occur at random time intervals.

The first step is to examine the varying seasonality indicators, including sinusoids and climate based moving averages (as defined in Section 1.8.14, page 39), to determine which measure 'best' describes the annual cyclic fluctuations that occur within the SIDS series. Both predictive ability and complexity of the variable are taken into account when defining the 'best' measure and forming a baseline model. Full details of this seasonality analysis follow in Section 3.2.

Each climatic variable is then examined individually at a univariate level for any potential relationship with SIDS, over and above the baseline model of an intercept and seasonality measure. The main aim at this stage of analysis is to eliminate variables that are not significantly related to SIDS. Climatic variables that are found to be related to the incidence of SIDS, over and above seasonality, will then be incorporated into a multivariate model relating SIDS to climate. This multivariate analysis is completed in Chapter 4, where principal component analysis is utilised to eliminate collinearity issues arising from the strongly interrelated climate measures. Potential problems due to serial correlation will be dealt with as the final step of the model building process in Chapter 4 by, when necessary, incorporating autoregressive terms into the model.

All analysis of the daily SIDS data is completed using logistic regression methods. The binary outcomes are days where at least one SIDS death occurred, and days with no SIDS deaths. Dichotomising the daily SIDS outcomes in this way incurred a minimal information loss: two of the 1827 days in Period 1 had two SIDS recorded (0.1%) and in Period 2, twenty of the 6209 days had two SIDS recorded (0.3%). Only one day in the complete study period had three SIDS deaths (the maximum number recorded), which occurred in September 1995. Full details of statistical methodology is given in the following section, with an overview of significant climatic relationships identified presented in Section 3.4. A complete description of all significant

results found using logistic regression modelling is presented in Appendix C.

3.1 Logistic Regression

The overall aim of this study is to locate the best fitting, yet meteorologically sensible model to delineate the relationship between SIDS and one or more climatic variables. This involves finding a balance between model complexity and interpretability. As with any analysis where the aim is to relate a response variable to a set of independent variables, regression methods are an important tool.

Logistic regression is a standard regression technique employed to model binary response data. It has been widely applied in medical and epidemiological fields, such as the development of post-operative liver failure (Kokudo et al., 2002), presence or absence of domestic violence (Jewkes et al., 2002), insemination status of malaria control mosquitoes (Okanda et al., 2002) and occurrence of childhood cancer (Okcu et al., 2002).

The general principles used in linear regression modelling are applied in logistic regression. A brief outline of the model and estimation methods is described below. Detailed coverage of logistic regression can be found in many textbooks, for example Hosmer & Lemeshow (2000) or Harrell (2001).

3.1.1 The Logistic Regression Model

Let y_i denote the ith response variable: $y_i = 0$ or 1. In this study this corresponds to

$$y_i = \begin{cases} 0 & \text{days with no SIDS deaths} \\ 1 & \text{days where at least one SIDS occurred} \end{cases} \tag{3.1}$$

with $i = 1, \ldots, n$ (n = total number of days within each study period). Let $\mathbf{y}$ represent the vector of responses $[y_1, y_2, \ldots, y_n]'$, let $\mathbf{X}$ denote the matrix of climatic covariates, $\mathbf{x}_i = [1, x_{1i}, x_{2i}, \ldots, x_{ki}]'$, where there are a total of k (independent) covariates.

The logistic regression model is then defined in terms of the probability that $y_i = 1$, given $\mathbf{x}_i$, with the specific form given by

$$\pi(\mathbf{x}_i) = E(y_i|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})}, \tag{3.2}$$

where $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_k]'$, the vector of regression coefficients. Equation 3.2 is the conditional mean of the logistic regression model, and as $\pi(\mathbf{x}_i)$ is a probability, it is formulated to avoid fitting probabilities outside the range of zero to one. The conditional distribution of the response variable is

$$y_i|\mathbf{x}_i \sim \text{Bin}[n, \pi(\mathbf{x}_i)]. \tag{3.3}$$

That is, conditional on $\mathbf{x}_i$, y_i is distributed Binomially with probability $\pi(\mathbf{x}_i)$.

The logit transformation of $\pi(\mathbf{x}_i)$ is given in equation 3.4 below,

$$\text{logit}[\pi(\mathbf{x}_i)] = \log\left[\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right] = \mathbf{x}_i\boldsymbol{\beta}. \tag{3.4}$$

This transformation is the canonical link function for logistic regression and it describes the relationship between $\pi(\mathbf{x}_i)$ and the covariates in the linear predictor. This monotonic functional form of the probability, given by the conditional mean, implies that $\text{logit}[\pi(\mathbf{x}_i)]$ is linear in $\mathbf{x}$. This is important, as the transformation results in the model having many covetable properties of the standard linear regression model.

3.1.2 Fitting the Logistic Regression Model

The likelihood of the logistic regression model is defined as:

$$l(\beta) = \prod_{i=1}^{n} \pi(\mathbf{x}_i)^{y_i}[1 - \pi(\mathbf{x}_i)]^{1-y_i}. \tag{3.5}$$

Maximum likelihood estimation (MLE) (Dobson, 2002) is used to calculate the parameter estimates in the logistic regression model. The MLE procedure involves maximizing the likelihood function given in equation 3.5 with respect to the parameters, β. The parameters that maximize $l(\beta)$ are then taken as the parameter estimates. They are denoted by $\hat{\beta}$, and defined as the maximum likelihood estimates (MLEs). Computationally, it is easier to maximise the log of the likelihood (defined as the log-likelihood). This is given by

$$L(\beta) = \log[l(\beta)] = \sum_{i=1}^{n} y_i \log[\pi(\mathbf{x}_i)] + (1 - y_i)\log[1 - \pi(x_i)] \tag{3.6}$$

$$= \sum_{y=1} \mathbf{x}_i\beta - \sum_{i=1}^{n} \log[1 + \exp(\mathbf{x}_i\beta)]. \tag{3.7}$$

Hence, parameter estimates are the values of β which maximize $L(\beta)$. They are computed numerically using an iterative weighted least squares procedure (Dobson, 2002).

3.1.3 Interpretation of the Coefficients of the Logistic Regression Model

After fitting a logistic regression model, which has an adequate fit, with covariates which are either statistically or contextually significant, the next step is to interpret the model. This is best achieved by examining the estimated model coefficients and making practical inferences, based on an odds ratio (OR). This helps assess the strength of the relationship between each independent covariate and the dependent variable, SIDS.

For the model given in equation 3.4, the odds of $y_i = 1$, that is, the odds of at least one SIDS occurring, for a unit increase in the jth covariate x_j, holding all other covariates fixed, is:

$$\widehat{OR}_i = \exp(\hat{\beta}_i). \tag{3.8}$$

The corresponding $100(1 - \alpha)\%$ confidence interval is given by

$$\exp[\hat{\beta}_i \pm z_{\frac{\alpha}{2}} \widehat{SE}(\hat{\beta}_i)] \tag{3.9}$$

where $\widehat{SE}(\hat{\beta}_i)$ is the estimated standard error of $\hat{\beta}_i$, and $z_{\frac{\alpha}{2}}$ is the normalised z statistic, corresponding to a $1 - \frac{\alpha}{2}$ significance level.

3.2 Statistics for Model Comparison

Five model fit statistics are defined below, which fall into two broad categories: log-likelihood based statistics and statistics based on varying definitions of model residuals. The statistics will be utilised as a basis for choosing between competing models to decide which model best fits the SIDS data.

3.2.1 Log-likelihood Based Statistics

The following three statistics are all based on the log-likelihood defined by equation 3.6, and evaluated at the fitted values $\hat{\beta}$.

Deviance Log-likelihood Statistic

Two forms of the deviance statistic are utilised in the model building process. The first, D, given in equation 3.10 below, is used to compare non-nested models:

$$D = -2L(\hat{\beta}) \tag{3.10}$$

where smaller values of D indicate the preferred model.

To determine the significance of an additional covariate to a model, a comparison is made between the values of the log-likelihood with and without the additional covariate in the model. The statistic is given by

$$\begin{aligned} G &= -2[L(\textit{for the model without variable}) - L(\textit{for the model with variable})] \\ &= D[\textit{for the model without variable}] - D[\textit{for the model with variable}]. \end{aligned} \tag{3.11}$$

Under the null hypothesis that β_i, the coefficient corresponding to the additional variable, equals zero, G is distributed χ^2 with one degree of freedom (assuming a large sample size).

Akaike's Information Criterion

Akaike's information criterion (AIC) is a form of a penalised likelihood statistic (Cameron & Trivedi, 1998). A penalty is added to the goodness-of-fit measure defined by the deviance statistic, D, in equation 3.10. This additional term is based on the number of parameters in the model, thereby the AIC statistic additively penalises for increased model complexity. The statistic is defined as

$$AIC = -2L(\hat{\beta}) + 2m \tag{3.12}$$

where m represents the number of parameters in the model. Like the deviance statistic D, smaller values of the AIC statistic denote the favoured model. The definition of the AIC statistic implies that the favoured model is assessed by a compromise between the number of parameters in the model and achieving a satisfactory model fit to the data.

Schwarz's Criterion

Schwarz's criterion (SC) is another form of a penalised likelihood function. As with the AIC, model adequacy is based on the log-likelihood and measured by the deviance statistic, D, in equation 3.10. The penalty imposed on this statistic is again, an addition of the number of parameters (m) in the model, though in this case, Schwarz's criterion scales m by the log of the sample size, n.

$$SC = -2L(\hat{\beta}) + m\log(n) \tag{3.13}$$

As previously, a smaller value of the statistic indicates the preferred model. If the sample size is greater than seven, the SC statistic will tend to prefer less complex models (fewer parameters) than those chosen by the AIC statistic (Hosmer & Lemeshow, 2000). The SC statistic is also widely known as the Bayesian information criterion, or BIC (McLachlan & Peel, 2000).

3.2.2 Residual Based Statistics

In regression analyses, residuals are defined as the difference between the outcome variable at an observed value and the predicted value of the outcome variable, as predicted by the regression model. With logistic

regression, several formulations of residuals are used to measure this deviation between the observed and fitted values.

In logistic regression, the fitted values are calculated for each distinct covariate pattern, rather than for each of the n observations. Following the notation of Hosmer & Lemeshow (2000), the fitted value is defined in equation 3.14 below. Let J represent the number of distinct covariate patterns, $\mathbf{x}_j$, observed in $\mathbf{X}$, the covariate matrix. Denote the number of observations with each covariate pattern, $\mathbf{x} = \mathbf{x}_j$ by m_j, $j = 1, 2, \ldots, J$. For example, m_1 is the number of observations with the covariate pattern $\mathbf{x}_1$. Among the m_j subjects with $\mathbf{x} = \mathbf{x}_j$, let y_j represent the total number of positive responses, that is, $y = 1$. The fitted value for the jth distinct covariate pattern, $\hat{y}_j$, is then given by

$$\hat{y}_j = m_j \hat{\pi}_j = m_j \frac{\exp(\mathbf{x}_j \hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{x}_j \hat{\boldsymbol{\beta}})}. \tag{3.14}$$

Note that when at least one of the covariates in the model is a continuous covariate, it is assumed that $J \approx n$.

Pearson Chi-squared Statistic

Utilising the fitted value in equation 3.14 above, the Pearson residual for the jth covariate pattern is defined as:

$$r(y_j, \hat{\pi}_j) = \frac{(y_j - \hat{y}_j)}{\sqrt{\hat{y}_j(1 - \hat{\pi}_j)}} = \frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}}. \tag{3.15}$$

The Pearson χ^2 statistic, P, is the resulting summary statistic formed by summing the Pearson residuals over all J covariate patterns, and is given by

$$P = \sum_{j=1}^{J} r(y_j, \hat{\pi}_j)^2. \tag{3.16}$$

The preferred model is indicated by lower values of P.

Hosmer-Lemeshow Statistic

The Hosmer-Lemeshow goodness-of-fit statistic (HL) was developed specifically for logistic regression models (Hosmer & Lemeshow, 1980; Lemeshow & Hosmer, 1982). It is based on a decile partitioning of the predicted probabilities, and then calculated via a standard χ^2 statistic comparing the mean predicted probability with the observed outcome, in each decile.

The notation of Hosmer & Lemeshow (2000) is used to define the HL statistic. Assume the number of distinct covariate patterns is equal to the sample size, that is $J = n$. There then exist n distinct predicted probabilities $\hat{\pi}_i$, $i = 1, 2, \ldots, n$. Let $\widehat{\boldsymbol{\Pi}}$ be the vector of predicted probabilities, $\widehat{\boldsymbol{\Pi}} = [\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_n]'$. The vector $\widehat{\boldsymbol{\Pi}}'$ is created by ranking $\widehat{\boldsymbol{\Pi}}$ in ascending order such that $\hat{\pi}'_1$ corresponds to the smallest $\hat{\pi}_i$ value, and $\hat{\pi}'_n$ to the largest $\widehat{\boldsymbol{\Pi}}$ entry. This ranked vector, $\widehat{\boldsymbol{\Pi}}'$, is then the basis of the grouping strategy used to calculate the HL statistic: partition $\widehat{\boldsymbol{\Pi}}'$ into ten groups corresponding to the deciles of $\widehat{\boldsymbol{\Pi}}'$. Both the observed and expected frequencies can then be calculated for each decile group.

Let c_k denote the number of distinct covariate patterns in the k^{th} decile group. The total number of positive responses among the c_k distinct covariate patterns, o_k, is then

$$o_k = \sum_{j=1}^{c_k} y_j,$$

where o_k corresponds to the total number of positive responses in the k^{th} group. The mean predicted probability, $\bar{\pi}_k$ is then given by

$$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n_k'}.$$

The Hosmer-Lemeshow statistic is then defined as

$$HL = \sum_{k=1}^{10} \frac{(o_k - N_k' \bar{\pi}_k)^2}{n_k' \bar{\pi}_k (1 - \bar{\pi}_k)} \tag{3.17}$$

where n_k' = the total number of subjects in the k^{th} group.
As with the previously defined statistics, a smaller value of the HL statistic denotes the preferred model.

3.3 Statistical Modelling Methods

Due to the exploratory nature of the analysis, a large number of meteorological measurements are examined in relation to SIDS. The measures (fully described in Chapter 1.8) essentially fall into six categories:

1. **Seasonality measures**,
2. **Daily summary measures** including minimum, mean and maximum measures for the day of interest and up to a fourteen day lag,
3. **Within day effects** measured by standard deviation, range and various measures of absolute hourly change, for the day of interest and up to a fourteen day lag,
4. **Between day effects** including the difference in daily measures on both consecutive days, and between the day of interest and a lag of up to fourteen days,
5. **Average day effects** examining the mean over past days, of up to a fortnight,
6. **Between average day effects** comparing the difference between the mean daily measure on the day of interest and a past average.

These variables are described fully in Section 1.8.2.

Logistic regression analysis was performed using SAS (*SAS (R) Proprietary Software Release (8.01.01)*) through the LOGISTIC and GENMOD procedures. Odds ratios are utilised throughout the modelling process to aid in the understanding and interpretation of the meteorological models. Period 1 (1968—1972) and Period 2 (1973—1989) are examined separately. The results are reported in Section 3.4.1 and 3.4.2 for each period respectively, with a detailed discussion given in Appendix C. A $\alpha = 0.05$ significance level is used throughout the analysis.

3.3.1 Seasonality

As discussed in Chapter 1, the seasonal pattern of SIDS occurrence over time has been noted for well over a century (Wakley, 1855). Many methods have been employed in the published literature to measure this cyclic variation. These methods essentially fall into three categories: sines and cosines (Douglas et al., 1998; Campbell et al., 2001), indicator variables (Campbell, 1989) and monthly summaries of various temperature measures (McGlashan & Grice, 1983; Schluter et al., 1998). These seasonality measures appear to have been arbitrarily chosen, with little justification, or validation.

This section of analysis methodically examines a multitude of measures of seasonality. The measures include those used in the literature, along with newly derived variables based on climatic measures other than temperature, which is the most common climate variable used in the literature to describe the seasonal pattern of SIDS deaths. Climatic variables examined as potential seasonality covariates include temperature, wind speed, wind velocity (in the four defined directions north, south, east and west), humidity, pressure, dewpoint, radiation, sunshine and rainfall.

The goodness-of-fit statistics defined in Section 3.2 are utilised to decide the best measure of the underlying seasonal distribution of the incidence of SIDS. When comparing the various climatic seasonal summaries, only D, P and HL are examined, as the three likelihood based methods (D, AIC and SC) will rank the climatic candidates in the same order. This is a result of the climatic models all having the same number of parameters and sample size, therefore the penalty added in both the AIC and SC is constant across all the climatic models.

A crude two step ranking procedure is applied to decide which candidate seasonality measure best describes the annual cyclic variation within the SIDS series:

> **Step 1**: The first step involves examining the various climatic measures and finding the best model among those. This is achieved by ranking the preferred models defined by each goodness of fit statistic, then summing the ranks for each variable. The variable corresponding to the model with the lowest rank-sum is then taken as the best climatic measure of seasonality.
>
> **Step 2**: The second step involves the same ranking procedure, this time comparing the best climatic measure found in step 1 against the first and second order sinusoids and the month indicator variable. The variable now corresponding to the model with the lowest rank-sum is chosen as the best measure of seasonality for the Canterbury SIDS data on a daily level.

Results from the seasonality analysis are reported in Section 3.4.1 and 3.4.2 for each period.

3.3.2 Daily Climatic Measures

Candidate climate covariates that are examined in association with SIDS incidence, over and above seasonality, include temperature, wind direction, wind speed, wind velocity and wind chill, humidity, pressure, rainfall, sunshine, solar radiation and dewpoint. Except for wind direction, solar radiation and sunshine, these variables are continuous and have measures defined for all the day, within day, between day and average variables. Table 1.3 (page 23), gives the full list of climate variables, and their corresponding abbreviations, used throughout this study.

Serial correlation between the seasonality indicator and candidate climatic covariates is a potential problem, as many of the meteorological variables have their own underlying seasonal pattern. Therefore correlations between all climatic measures and the seasonality indicator will be examined. In the case of a strong association between seasonality and a climatic measure, the climatic measure will be 'deseasoned'. This is achieved by subtracting the seasonal component from the climatic measure under investigation. The deseasoned variable will then be examined in relation to SIDS.

In addition to being seasonally dependent, the climatic measures have the potential to be strongly associated with each other. Therefore initial analysis will involve creating individual regressions for each measure, of the form *baseline model + climate variable*, and examining the significance of the resulting model. In the final stage of the model building process, a multivariate regression model will be created in such a way that the problem of interrelated covariates is resolved. This section of analysis is presented in Chapter 4.

	D	AIC	SC	P	HL	Rank sum
1. **Sinusoids**						
$\sin(\frac{2\pi t}{365}) + \cos(\frac{2\pi t}{365})$	638.26	644.26	660.79	1835.05	6.06	14
2. **Sinusoids with second harmonics**						
$\sin(\frac{2\pi t}{365}) + \cos(\frac{2\pi t}{365}) + \sin(\frac{4\pi t}{365}) + \cos(\frac{4\pi t}{365})$	636.59	646.59	674.14	1822.10	5.78	12
3. **Month indicator variable**						
month(i), $i = 1, 2, \ldots, 12$	627.91	651.91	718.04	1827.00		16
4. **Retrospective climatic average**						
$Temp_{meanMA30}$	631.84	635.84	646.82	1800.69	12.62	5

Table 3.1: Goodness of fit statistics for selected seasonality models, Period 1.

Parameter	Estimate (SE)	Odds ratio (95% CI)
Intercept	-2.006 (0.119)	
$Temp_{meanMA30}$	-0.078 (0.011)	0.925 (0.905, 0.944)

Table 3.2: Model estimates for the seasonality model, Period 1.

3.4 Results from Logistic Regression Modelling

As the primary focus of this analysis section is data reduction, only a brief overview of covariate trends and relationships is presented. Appendix C gives a full list of the climate variables identified as being significantly related to the incidence of SIDS (at $\alpha = 0.05$), over and above the seasonality component. These are presented in Tables C.1 to C.29, which detail parameter estimates, associated odds ratios and goodness-of-fit statistics. Appendix C also provides a detailed discussion covering selection between competing variables, and parameter interpretations.

3.4.1 Period 1 (1968—1972)

Seasonality Measure

The resulting goodness-of-fit statistics for the models corresponding to four candidate seasonality variables are presented in Table 3.1. The first step in the ranking process resulted in one climatic measure appearing best: $Temp_{meanMA30}$ (the mean daily temperature averaged over the past thirty days). Within this first ranking step $Dew_{meanMA30}$ had a final rank sum of 10, while the rank sum of $Temp_{meanMA30}$ was 9. Although $Dew_{meanMA30}$ was closely ranked to $Temp_{meanMA30}$, it was decided to retain only the temperature summary as it is a more traditional measure, that has been previously applied to this data set (Schluter et al., 1998).

The rank sums for the second stage of the ranking procedure are presented in the final column of Table 3.1. These show that the temperature measure outperforms the more contrived and mathematically based measures when modelling seasonality in daily SIDS occurrences in Period 1 (1968—1972) of the study period. The variable $Temp_{meanMA30}$ will therefore be retained as the daily seasonality measure for Period 1, and used in further analysis of the relationship between SIDS and climate.

Full parameter estimates, standard errors and odds ratios for the model combining $Temp_{meanMA30}$ are presented in Table 3.2. This model indicates that SIDS risk decreased with every 1°C increase in $Temp_{meanMA30}$. Relating this in terms of absolute SIDS risk, when $Temp_{meanMA30} = 0$°C, the odds of at least one SIDS occurring was 0.134, just over 1 in 8 days. At $Temp_{meanMA30} = 5$°C this risk decreased to 0.091, one in eleven days, while when $Temp_{meanMA30} = 10$°C the risk of at least one SIDS decreased again

to 0.061, approximately one in seventeen days.

It is interesting to note that overall the retrospective moving averages of climatic summary variables appear to capture the underlying seasonal pattern in SIDS incidence better than the monthly summary variables in Period 1. Of the three daily summaries (minimum, mean and maximum), averaged over the seasonal period, no one outperforms the others as a seasonal measure across all the climatic variables. In comparing the first order sinusoid against the sinusoid with second harmonics (1. and 2. in Table 3.1), the more complex seasonality measure performs better when measured by both the HL and D statistics. Yet, when a penalty is added for model complexity, the first order sinusoid is preferred using both the AIC and SC.

Climate Measures

Temperature: Temperature measures significantly associated with the incidence of SIDS in Period 1, after adjusting for seasonality, essentially fell into three groups: (1) An increase in temperature on the day of interest ($day0$) related to an increase in the risk of SIDS; (2) An increase in the variability of temperature a week ago, also corresponded to an increased SIDS risk; (3) SIDS risk was found to increase when the temperature on $day0$ was cooler than it had been over the past two weeks (Table C.1). For example, a five degree increase in $Temp_{diffday0-14}$ from 0°C to 5°C, results in the odds of at least one SIDS occurring increasing from 0.04 to 0.06 (holding $Temp_{meanMA30}$ fixed at its mean value of 11.44).

Wind Direction: Measures of wind direction fall into two classes: categorical measures relating to the predominant wind direction, and various continuous measures based around the absolute hourly change in wind direction. Compared with the 'mixed' wind direction category (where no one direction predominated), a predominantly south-westerly wind on $day-1$ (the day before $day0$), or a week prior, related to a decreased risk of SIDS. In contrast, compared with the 'mixed' category, a calm day (no wind) led to an increased SIDS risk eight days later. A change in the variability of the wind direction over the past few days was also related to the SIDS incidence (Tables C.2 and C.3).

Wind Speed: The addition of most wind speed variables to the baseline model of intercept and seasonality resulted in seasonality no longer remaining significantly related to SIDS, although wind speed was not associated with seasonality. Generally, an increase in either wind speed itself, or the variability in wind speed up to one week prior to $day0$ corresponded to a decrease in the risk of SIDS. Also, a greater SIDS risk occurred when the wind speed on $day0$ was stronger than it had been in the past (Table C.4).

Wind Velocity — North: The strength, and variability, in the northern component of wind velocity on both $day0$ and eight days prior were related to SIDS incidence. A variety of between day measures ranging from $day0$ to a week prior were also associated with the incidence of SIDS, over and above seasonality. An increase in northern wind velocity on $day0$ from the average over the past few days related to an increased risk of SIDS (Table C.5). This increase in SIDS risk, with increasing $North_{diffday0-14}$ is illustrated in Figure 3.1.

Wind Velocity — South: Two sets of southern wind velocity variables were associated with the incidence of SIDS, after accounting for seasonality: (1) An increase in the strength of wind from the south a week prior to $day0$ corresponded to a decreased risk of SIDS; (2) In contrast, an increase in the variability of the southern wind velocity on $day0$ from what it was five days prior led to an increased SIDS risk (Table C.6).

Wind Velocity — East: Of all the daily variables measuring the easterly component of wind velocity, only three were associated with SIDS, over and above seasonality. All related to a measure of the variability in eastern wind velocity, and were associated with both increased and decreased SIDS risk (Table C.7).

Wind Velocity — West: Variables measuring the western component of wind velocity that were significantly related to SIDS rates, over and above seasonality, cover three broad constructs; an increase in

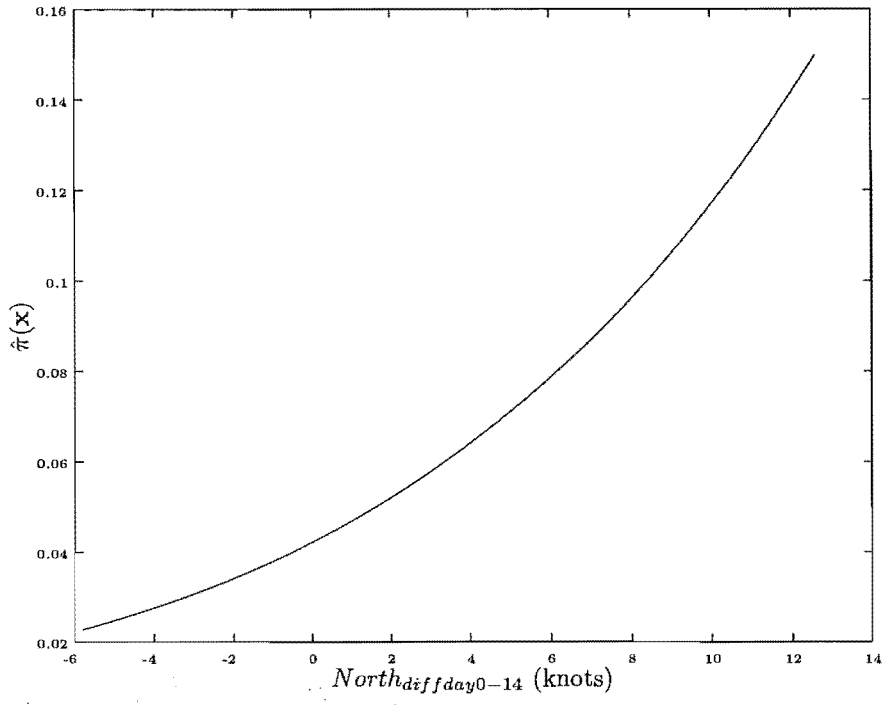


Figure 3.1: Risk of recording at least one SIDS per day, with respect to $North_{diffday0-14}$, Period 1.

the strength of the westerly wind velocity resulted in a decreased risk of SIDS two and seven days later; similarly, an increased SIDS risk was associated with increased variability in western wind velocity up to a week prior to *day*0; an association was found between SIDS incidence and the average wind velocity from the west over the past few days (Table C.8). For example, an increase in the average strength of the western component of wind velocity over the three days prior to *day*0 ($West_{meanday-1to-3}$) from 10 to 20 knots, results in the odds of at least one SIDS occurring decreasing from approximately one in 65 days to one in 350 days (holding $Temp_{meanMA30}$ fixed).

Wind Chill: An increase in the wind chill index on *day*0 corresponded to an increased risk of SIDS. In contrast, an increase in the variability of wind chill related to a decreased SIDS risk, after accounting for the seasonal component in the SIDS series (Table C.9).

Relative Humidity: Out of all the candidate humidity measures examined in relation to SIDS on a daily time scale, only three variables were significantly associated with the incidence of SIDS. The three covariates were all measuring the variability in humidity, up to five days prior to *day*0, where an increase in any of the variables inferred a subsequent increase in SIDS risk (Table C.10). This trend is illustrated for $Humid_{stdAHCday-4-5}$ in Figure 3.2.

Pressure: A number of measures of the variability in pressure were found to be significantly associated with SIDS, over and above seasonality (Table C.11). Essentially, these models are interpreted as: the more variable the pressure, the smaller the risk of SIDS.

Rainfall: The rainfall on *day* − 1 was important in terms of SIDS. An increase in the intensity, or the variability, of rainfall on that day related to a decreased risk of SIDS. Table 3.3 gives parameter estimates and corresponding odds ratios for the Period 1 rainfall variables that were selected for inclusion in further analysis. This highlights the negative relationship between rainfall and the incidence of SIDS in Period 1. Full rainfall details appear in Table C.12.

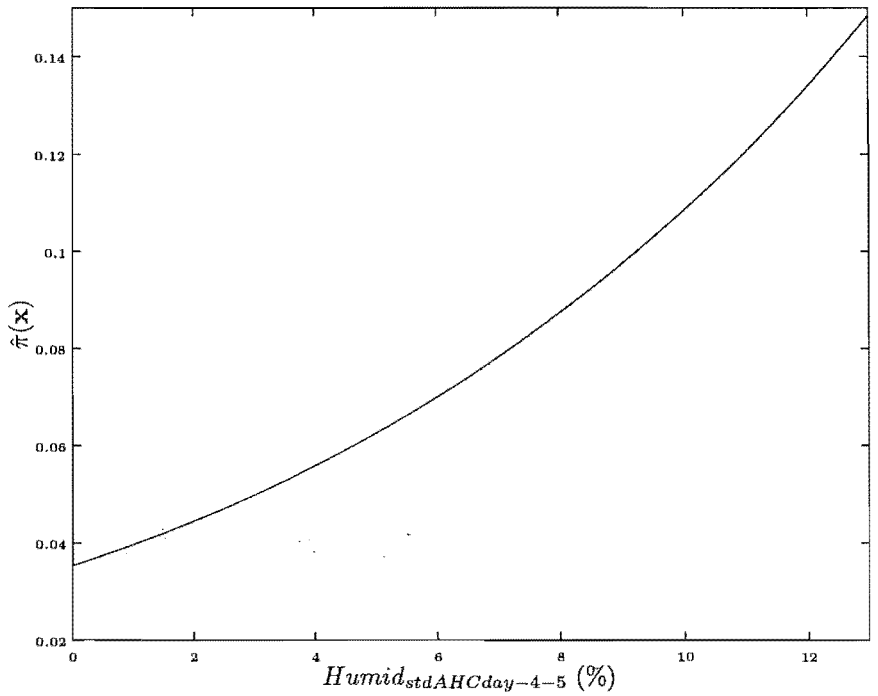


Figure 3.2: Risk of recording at least one SIDS per day, with respect to $Humid_{stdAHC-4-5}$, Period 1.

Variable	Estimate	Odds ratio (95% CI)
$Rain_{maxday-1}$	-0.368	0.692 (0.473, 1.013)
$Rain_{stdday-1}$	-1.558	0.211 (0.046, 0.969)
$Rain_{stdday0-1}$	-1.134	0.322 (0.104, 0.999)

Table 3.3: Rainfall variables significantly related to the incidence of SIDS in Period 1, and retained for futher analysis.

Sunshine: Only two of the sunshine variables examined were found to be significantly associated with the incidence of SIDS, over and above seasonality. Both variables were between day effects and implied that SIDS risk increased with an increasing difference in the total number of sunshine hours between days (Table C.13).

Solar Radiation: After adjustment for seasonality, no relationship between solar radiation and SIDS incidence were identified.

Dewpoint: A variety of dewpoint covariates were associated with SIDS, over and above the seasonal component. An increase in the dewpoint resulted in a corresponding increase in the risk of SIDS four days later. Similarly, of the many variables measuring the between day variability of dewpoint, the majority reflected an increased variability in dewpoint implying an increase in the risk of SIDS. Situations when the dewpoint on $day0$ is higher than it had averaged over the past fourteen days related to an increased risk of SIDS (Table C.14). Specifically, an increase in $Dew_{diffday0-14}$, from $-5°C$ to $5°C$, leads to a corresponding increase in the odds of recording at least one SIDS from 0.03 to 0.07 per day.

Logistic Regression with Multiple Predictors: Table C.15 in Appendix C shows the results from a logistic regression with multiple predictors. All 125 candidate climate variables were initially included in this model, and a backwards elimination procedure used to give the best fit model. Significance was taken as $\alpha \leqslant 0.05$. The resultant model (Table C.15 shows a multi-covariate relationship between SIDS and climate and contains 25 climate-based regressors. Interpretation of this model in terms of a risk profile of SIDS incidence is impractical. Principal component regression (PCR, see Chapter 4 for details) is therefore used to find a reduced set of constructs to summarise the SIDS climate relationship. PCR enables the complete climatic structure to be examined in relation to SIDS, and potentially highlight a more detailed, and potentially more interpretable climate SIDS model.

3.4.2 Period 2 (1973—1989)

Seasonality Measure

Table 3.4 presents the goodness of fit statistics for the models corresponding to four candidate seasonality measures for Period 2. The initial ranking procedure (as described in Section 3.3.1) examined the models resulting from measuring seasonality via retrospective and calendar monthly averages of climatic covariates. The model based on $Temp_{minMA30}$ (the minimum daily temperature averaged over the past thirty days) was considered best out of all the climatic measures. This model was then compared to the sinusoid models and the indicator variable model. With a rank sum of ten, the model corresponding to seasonality measured by a first order sinusoid function was considered the best measure of seasonality to model the annual periodicity in daily SIDS occurrences in Period 2 (1973—1989). This variable will be utilised as the daily seasonality measure for Period 2 in further analysis.

Full parameter estimates, standard errors and odds ratios for the model containing first order sinusoids are presented in Table 3.5. To uncouple the two parts of the seasonal sinusoid function is misleading, therefore interpreting the full model in terms of absolute SIDS risk showed that the SIDS risk was 0.042 in summer, whereas in winter this probability increased nearly threefold to 0.117.

In comparing the various measures of seasonality, it is interesting to note that when comparing the monthly averages against the retrospective moving averages, the deviance based statistics prefer the moving average based models, whereas the HL residual based statistic prefers the monthly average measures. Differentiating between the different averages was not possible using the Pearson residual based statistic. No obvious differences were identified between seasonality measured via retrospective moving averages of climate variables, or seasonality measured as monthly averages of the climate variables. Similarly, no clear

	D	AIC	SC	P	HL	Rank sum
1. **Sinusoids**						
$\sin(\frac{2\pi t}{365}) + \cos(\frac{2\pi t}{365})$	3250.33	3256.33	3276.53	6192.44	7.17	10
2. **Sinusoids with second harmonics**						
$\sin(\frac{2\pi t}{365}) + \cos(\frac{2\pi t}{365}) + \sin(\frac{4\pi t}{365}) + \cos(\frac{4\pi t}{365})$	3241.27	3251.27	3284.94	6232.51	9.11	12
3. **Month indicator variable**						
month(i), $i = 1, 2, \ldots, 12$	3232.57	3256.57	3337.37	6209.00		16
4. **Retrospective climatic average**						
$Temp_{minMA30}$	3252.39	3256.39	3269.84	6178.93	10.44	12

Table 3.4: Goodness of fit statistics for selected seasonality models, Period 2.

Parameter	Estimate (SE)	Odds ratio (95% CI)
Intercept	-2.574 (0.051)	
$\sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)
$\cos(\frac{2\pi t}{365})$	-0.513 (0.071)	0.599 (0.521, 0.689)

Table 3.5: Model estimates for the first order sinusoid model, Period 2.

differences were apparent between averages of minimum, mean or maximum daily measures. In comparing the first order sinusoid function with the sinusoid function with second harmonics, the more complex model was preferred by the AIC and D statistics, whereas the remaining three goodness-of-fit statistics preferred the simpler functional form.

Climate Measures

Temperature: The only temperature variables found to be associated with the incidence of SIDS, over and above seasonality, measured the variation in temperature, up to a fortnight ago. Essentially, an increase in the variation in temperature corresponded to an increased risk of SIDS (Table C.16). This relationship is illustrated graphically in Figure 3.3, which highlights the positive trend of increasing $Temp_{meanday0-14}$ corresponding to a predicted increase in the risk of SIDS.

Wind Direction: Three lagged variables relating to the predominant wind direction were found to be significantly associated with SIDS, over and above seasonality (Table C.17). Taking the 'mixed' category as the reference category, SIDS risk decreased with north-easterly wind three days before $day0$. In contrast, an easterly wind four days before $day0$, or a calm day or north-westerly wind a week prior to $day0$, were all related to an increased risk of SIDS. Table C.18 presents details of significant AHC wind direction variables.

Wind Speed: An increase in the strength, and variability, in the wind speed on $day0$ were related a decreased risk of SIDS. For example, an increase in the maximum wind speed on $day0$ ($WindS_{maxday0}$) from 10 to 20 knots decreases the odds of recording at least one SIDS per day from 0.083 to 0.064 (holding the seasonal sinusoid function fixed at $\sin(\frac{2\pi t}{365}) = -0.866$ and $\cos(\frac{2\pi t}{365}) = 0.50$, approximate values for spring). A variety of between day measures, ranging from $day0$ to a week before $day0$, were also associated with the incidence of SIDS, over and above seasonality (Table C.19).

Wind Velocity — North: The strength, and variability, of the northern component of wind velocity both a week, and a fortnight prior to the day of interest were related to SIDS incidence. An increase in between day measures capturing the difference in northern wind velocity from $day0$ to a week or fortnight prior corresponded to an increase in the risk of SIDS (Table C.20).

Wind Velocity — South: Significant relationships identified between southern wind velocity measures

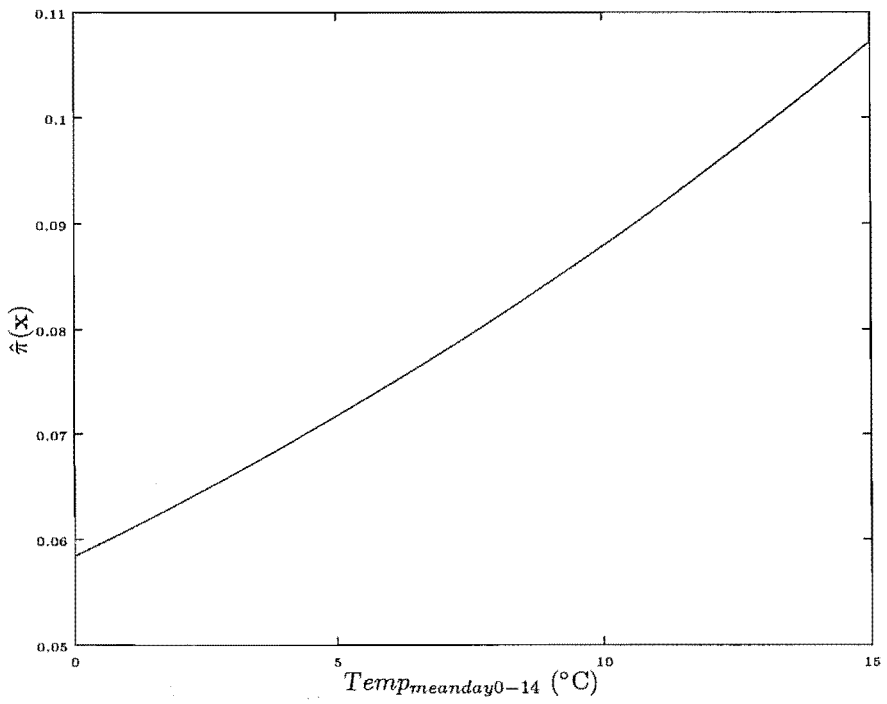


Figure 3.3: Risk of recording at least one SIDS per day, with respect to $Temp_{meanday0-14}$, Period 2.

and SIDS essentially fell into three groups: (1) An increase in the strength, and variability, of the southern wind velocity three or four days before the day of interest related to an increased risk of SIDS; (2) Between day measures examining the difference in southern wind velocity on consecutive days were also associated with the incidence of SIDS, over and above seasonality; (3) SIDS risk decreased with increased southern wind velocity, on average, a few days prior to *day*0 (Table C.21).

Wind Velocity — East: An increase in the eastern component of wind velocity corresponded to an increase in SIDS risk three days later, whereas, an increase in the variability in eastern wind velocity over the past couple of days related to a decreased risk of SIDS. Table 3.6 gives parameter estimates and corresponding odds ratios for the Period 2 easterly variables that were selected for inclusion in further analysis. This highlights the both the positive and negative relationships between varying easterly wind velocity variables and the incidence of SIDS in Period 2.

Wind Velocity — West: A decreased SIDS risk was associated with both increased strength and variability in the western component of wind velocity three days prior. Various between day measures were also significantly related to the incidence of SIDS, after accounting for seasonality, and an increase in the mean western wind velocity over a few days prior to the day of interest corresponded to an increased risk of SIDS (Table C.23).

Wind Chill: After adjusting for seasonality, wind chill was found to have no significant relationship with the incidence of SIDS in Period 2.

Relative Humidity: The maximum relative humidity a fortnight before *day*0, and various measures of the variability in humidity, up to six days prior to *day*0, were all negatively associated with the incidence of SIDS. This is illustrated in Figure 3.4, which highlights a decreased risk of SIDS with increasing $Humid_{minday0-2}$.

Pressure: An increase in pressure up to a week prior to *day*0 led to an increase in the risk of SIDS.

Variable	Estimate	Odds ratio (95% CI)
$East^{d}_{meanday-3}$	0.038	1.038 (1.008, 1.069)
$East^{d}_{maxday-3}$	0.024	1.025 (1.007, 1.042)
$East^{d}_{stdday-3}$	0.076	1.079 (1.027, 1.134)
$East^{d}_{rangeday-3}$	0.026	1.026 (1.008, 1.045)
$East^{d}_{stdAHCday-3}$	0.141	1.151 (1.026, 1.292)
$East_{maxAHCday-1-2}$	-0.053	0.948 (0.911, 0.986)
$East_{stdAHCday-1-2}$	-0.155	0.857 (0.737, 0.997)

Table 3.6: Eastern wind velocity variables significantly related to the incidence of SIDS in Period 2, and retained for further analysis ('d' denotes variables that have been deseasoned).

Similarly an increase in the variability of pressure up to a week before *day*0 was associated with a significant increase in SIDS risk, over and above seasonality. A decreased risk of SIDS occurred on days where the pressure was lower than it had averaged over the past week (Table C.25).

Rainfall: An increase in rainfall intensity, or variability, on *day*0 indicated a corresponding increase in the risk of SIDS. Numerous variables capturing the difference in rainfall and various lags were positively associated with SIDS. For example, an increase in the difference between the maximum rainfall on *day*0 and $day-1$ ($Rain_{maxday0-1}$) from 0 to 10 mm, increases the odds of recording at least one SIDS from approximately 1 in 15 days to 1 in 8 days (holding remaining covariates fixed). In contrast, an increase in the average rainfall over the past fortnight related to a decreased SIDS risk. Full rainfall details appear in Table C.26.

Sunshine: After adjusting for seasonality, the total hours of sunshine on *day*0 is negatively associated with SIDS risk. In contrast, a sunny day three days prior to *day*0, increased the risk of SIDS. An increased risk of SIDS also occurred on days where there had been less sunshine than on average over the few days prior to *day*0 (Table C.27).

Solar Radiation: Similar associations were identified between SIDS and solar radiation, as found between SIDS and sunshine: radiation on *day*0 was negatively associated with SIDS incidence; and days where there was less radiation than on average over the past few days was associated with a decreased risk of SIDS (Table C.28).

Dewpoint: Increased dewpoint over the few days prior to *day*0, or increased variability in dewpoint on *day*0 were related to a decreased SIDS risk. Specifically, a three degree increase in $Dew_{stdAHCday0}$, from 0°C to 3°C, decreases the odds of recording at least one SIDS per day from 0.093 to 0.037. Various between day measures were also significantly related to the incidence of SIDS, after accounting for seasonality (Table C.29).

Logistic Regression with Multiple Predictors: Table C.30 in Appendix C shows the results from a logistic regression with multiple predictors. All 171 candidate climate variables were initially included in this model, and as described in the Period 1 results section, a backwards elimination procedure used to give the best fit model. The resultant model (Table C.30 again highlights the multi-covariate relationship between SIDS and climate, and containing 35 climate-based regressors. Interpretation of this model in terms of a risk profile of SIDS incidence is impractical. As in Period 1, principal component regression (PCR, see Chapter 4 for details)) is therefore used to find a reduced set of constructs to summarise the SIDS climate relationship in Period 2.

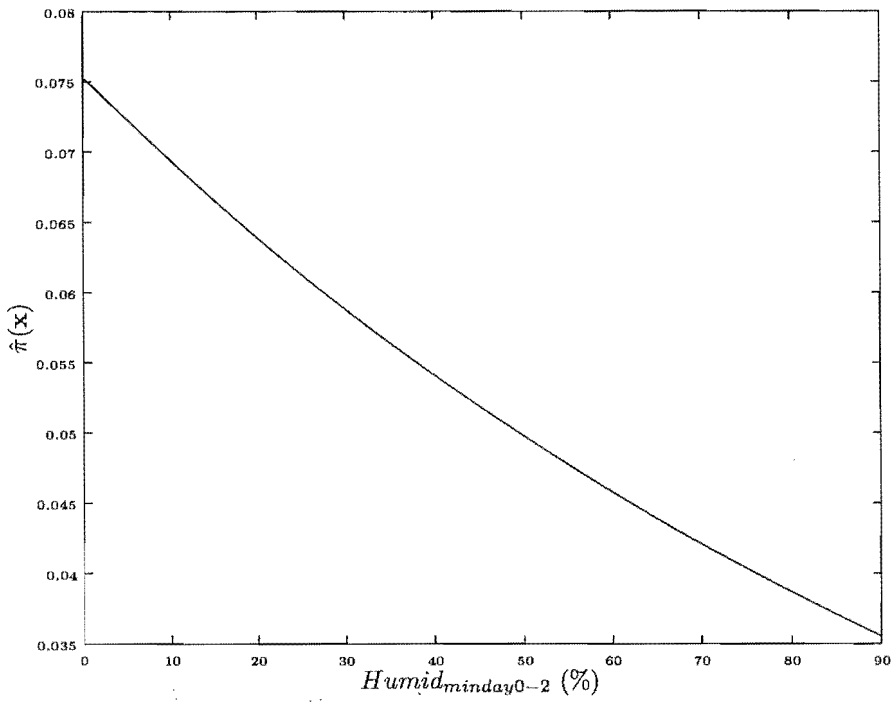


Figure 3.4: Risk of recording at least one SIDS per day, with respect to $Humid_{minday0-2}$, Period 2.

3.5 Summary and Conclusion

Logistic regression modelling of the daily relationship between the incidence of SIDS and climatic covariates identified various associations. The best indicator of seasonality in Period 1 was found to be $Temp_{meanMA30}$, a thirty day retrospective average of the mean temperature. In Period 2, the variable that best described the annual cyclic fluctuations in the SIDS series was a first order sinusoid, calculated using an additive combination of both sine and cosine functions with a period of 365 days.

The second stage of this section of analysis examined the significance of all daily climatic covariates individually in relation to SIDS incidence, after accounting for seasonality. As a data reduction step, this successfully decreased the number of climate variables from approximately 2500 to 125 distinct variables potentially associated with SIDS in Period 1, and a reduction to 171 distinct variables in Period 2.

From this initial analysis, the relationship between SIDS and climate in Period 1 appears to differ from that in Period 2. Obvious examples of this include solar radiation, which had no relationship with SIDS, after adjustment for seasonality in Period 1, compared to Period 2 where it was negatively associated with SIDS. Similarly, wind chill was not significantly associated with SIDS incidence in Period 2, yet in Period 1, both positive and negative associations were identified between SIDS and various wind chill measures.

This exploratory section of analysis involved investigating a multitude of climate variables. As a result of this, some of the significant associations identified may be Type I errors. Rather than using the Bonferroni method of adjustment for multiple comparisons (Cliff, 1987), which is generally accepted as being a conservative adjustment (Rothman & Greenland, 1998), all potential covariates are retained. Further examination of the relationship between climate and SIDS is developed in the next chapters, where a detailed climatic risk profile for the incidence of SIDS is constructed. Varying time scales and methodology are used and these identify, with more certainty, the underlying structure of the climatic dependency of SIDS incidence

in Canterbury.

Significant relationships identified in this univariate-type analysis must be interpreted with caution, as the weather on any day is constructed of intricate interrelationships and interdependencies between the individual climatic measures. Sections 3.4.1 and 3.4.2 give only a brief description of the general trends found as an overview of the SIDS—climate relationship. Logistic regression models with multiple predictors were constructed for both Period 1 and Period 2. Both models included an extensive representation of the available climate variables, and with 25 and 35 climate-based regressors in Period 1 and Period 2 respectively, practical interpretation of the model was not possible. Principal component regression is therefore used as a mechanism to model the daily relationship between SIDS and climate, with details presented in the next chapter.

Chapter 4

Principal Component Regression Analysis of Daily SIDS Counts

Chapter 3 initially examined approximately 2500 climate variables at a univariate level for any relationship with SIDS. A complete climatic profile of SIDS dependency will be constructed in this chapter using the 125 variables in Period 1, and 171 variables in Period 2, that were identified as being significantly related to the incidence of SIDS. The climate variables presented in Chapter 3, by nature of their definition, are naturally related. This is expected as all variables are measured at the same location, at the same time. In Christchurch, it is well known that a strong north-west wind in summer is associated with hot, dry days. In contrast, a southerly wind in winter is associated with rain and bitter temperatures. It is expected that on a day with a high total rainfall, sunshine hours and solar radiation measures will be less as a result of the cloud cover associated with rain. In turn, it is expected that, for example, the various pressure deviates will have some relationship as a result of deriving from the same original hourly data. Significant evidence of multicollinearity among the majority of the daily climate variables was identified via cross-correlations (Chatfield, 1996). Therefore a multiple logistic regression model cannot be developed from the raw climate variables as they violate the independent covariates requirement. To overcome this problem principal component regression will be utilised.

Principal component regression is a technique where the correlated covariates are replaced by uncorrelated principal components in the regression analysis. The principal components are, by definition, not correlated so multicollinearities do not exist between them, and the regression requirements are satisfied.

This chapter gives a brief overview of principal component analysis, and the extension to principal component regression with respect to the binary outcome situation presented by the daily SIDS data. Finally, results are presented for each period, for the multiple climatic components modelling.

4.1 Principal Component Analysis

Principal component analysis (PCA) is a data reduction method that aims to construct uncorrelated components $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_k$, which are formed from linear combinations of k variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$. As the components are not correlated, they are essentially capturing different dimensions of the original data. The components (known as principal components, or PCs) are arranged in descending order with respect to the amount of variation of the original variables each accounts for.

Many multivariate texts present detailed theory on PCA (for example Jolliffe (2002), Cliff (1987) or Harman (1976)); the following outline utilises the notation of Manly (1986).

Let $\mathbf{X}$ be the $n \times k$ matrix of climatic variables to be incorporated into the PCA, where n is the sample size for each period and k the number of climatic measures. The ith principal component ($i = 1, \ldots, k$) is given by the following combination of the variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$,

$$\mathbf{Z}_i = a_{i1}\mathbf{x}_1 + a_{i1}\mathbf{x}_2 + \cdots + a_{ik}\mathbf{x}_k \tag{4.1}$$

where

$$a_{i1}^2 + a_{i2}^2 + \cdots + a_{ik}^2 = 1. \tag{4.2}$$

The component loadings, a_{ij}, are calculated by maximizing the variance of $\mathbf{Z}_i$. Without the constraint in equation 4.2, simply increasing any of the component loadings leads to an increase in the variance of $\mathbf{Z}_i$.

Define $\mathbf{C}$ as the sample covariance matrix of $\mathbf{X}$, then principal component analysis simply involves calculating the eigenvalue (λ_i) — eigenvector ($\mathbf{e}_i$) pairs of $\mathbf{C}$. This results of the following relationships:

$$\begin{aligned} \text{Variance of } \mathbf{Z}_i &= \lambda_i \\ \text{Component loadings } \mathbf{a}_i &= \mathbf{e}_i \end{aligned} \tag{4.3}$$

To eliminate the possibility of one variable unjustifiably influencing the PCs, the variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$, are normalised to have zero mean and unit variance at the outset of the analysis.

To aid in the practical interpretation of the PCs, the matrices of PC loadings are often rotated. The general idea behind rotating is to have the original variables loading high on one component, and low on the others, while maintaining the orthogonal structure of the components. Those variables that correspond to high component loadings on a rotated component are then considered to define the structure of that particular component.

The most widely used rotation method is varimax rotation (Cliff, 1987), which uses an orthogonal component rotation. The following varimax criterion is maximised by rotation of the matrix of component loadings (a_{ij}) (Harman, 1976). The varimax criterion is given by

$$V = n\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij})^4 - \sum_{i=1}^{m}\left(\sum_{j=1}^{n}a_{ij}^2\right)^2 \tag{4.4}$$

where m is the number of components that are rotated.

As the primary aim of PCA is to generate a reduced set of uncorrelated components, which account for a large proportion of the variation in the original data set, a subset of size m of the k PCs are retained for rotation, interpretation and subsequent regression analysis. Many methods have been proposed to choose m, ranging from formal tests of hypotheses (Kim & Mueller, 1978), to subjective choices based on interpretability of the rotated components (Cliff, 1987). Three basic methods are utilised in this section of analysis, as the data will be reduced further in the PC regression procedure in Section 4.2. The methods for choosing m are:
Method 1: Retain only $\mathbf{Z}_i$s such that the proportion of variance explained by each component ($\text{var}(\mathbf{Z}_i)$) is greater than or equal to 0.01.
Method 2: Retain m components where m is the smallest number of $\mathbf{Z}_i$s that $\sum_{i=1}^{m}\text{var}(\mathbf{Z}_i) \geqslant 0.90$.
Method 3: Retain m $\mathbf{Z}_i$s, where m is defined as:

$$m = \frac{\text{total weighted variance}}{k}.$$

4.2 Principal Component Regression

Principal component regression (PC regression) is a method for dealing with a dataset containing strongly related variables. Instead of using the original (correlated) variables as covariates in a regression analysis, the principal components of the original data are calculated, and then these (uncorrelated) PCs are substituted as the independent variables for the regression.

In the case of linear regression, if all k PCs are incorporated into the analysis, the resultant regression model is the same as that obtained by least squares (Jolliffe, 2002). Jolliffe (2002) presents the theory for PC regression in the linear regression situation. This has been adapted for the case of the binary outcomes of SIDS or no SIDS days, as required for this section of analysis.

The logistic regression model was presented in Chapter 3, equation 3.2 (page 63). Let $\mathbf{X}^*$ denote the matrix of normalised climatic covariates, $\mathbf{x}_i^* = [x_{1i}^*, \ldots, x_{ki}^*]'$, where there are a total of k covariates. The logit transformation in terms of $\pi(\mathbf{X}^*)$ is then

$$\text{logit}[\pi(\mathbf{X}^*)] = \log[\frac{\pi(\mathbf{X}^*)}{1-\pi(\mathbf{X}^*)}] = \mathbf{X}^*\boldsymbol{\beta}. \tag{4.5}$$

The matrix analogue of equation 4.1, the PCs, is

$$\mathbf{Z} = \mathbf{X}^*\mathbf{A} \tag{4.6}$$

where the ith column of $\mathbf{Z}$ is $\mathbf{Z}_i$ as defined in equation 4.1, and $\mathbf{A}$ is the $(k \times k)$ matrix of component loadings. $\mathbf{A}$ is orthogonal, as it consists of normalised eigenvectors, derived from the sample covariance matrix, therefore

$$\mathbf{X}^*\boldsymbol{\beta} = \mathbf{X}^*\mathbf{A}\mathbf{A}'\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\gamma} \tag{4.7}$$

where $\boldsymbol{\gamma} = \mathbf{A}'\boldsymbol{\beta}$. Equation 4.5 then becomes

$$\text{logit}[\pi(\mathbf{X}^*)] = \mathbf{Z}\boldsymbol{\gamma} \tag{4.8}$$

which gives

$$\pi(\mathbf{X}^*) = \frac{\exp(\mathbf{Z}\boldsymbol{\gamma})}{1+\exp(\mathbf{Z}\boldsymbol{\gamma})}. \tag{4.9}$$

This is simply equation 4.5 with the climatic covariates substituted by their corresponding principal components in the logistic regression model. PC regression is also defined for the reduced model, with m components retained, as follows:

$$\pi(\mathbf{X}^*) = \frac{\exp(\mathbf{Z}_m\boldsymbol{\gamma}_m)}{1+\exp(\mathbf{Z}_m\boldsymbol{\gamma}_m)}. \tag{4.10}$$

where $\boldsymbol{\gamma}_m$ and $\mathbf{Z}_m$ are the corresponding subsets of $\boldsymbol{\gamma}$ and $\mathbf{Z}$ respectively.

The statistics defined in Section 3.2 (page 64) will be utilised to identify the best model.

4.3 Assessing the Fit of the Model

After completing analysis via PC regression, and having identified the best model to relate SIDS to climate components, the model needs to be examined for its appropriateness and adequacy in terms of how effectively it describes the incidence of SIDS. To completely assess the adequacy of the model both calculation of summary measures of overall model fit and examination of the contribution of the individual observations need to be considered.

4.3.1 Overall Model Fit

Goodness-of-fit Tests

Two of the statistics introduced for model comparison (Section 3.2, page 64) are utilised as a way of formally testing for the overall fit of the candidate model, namely Pearson's χ^2 statistic, P (equation 3.16, page 66), and the Hosmer-Lemeshow statistic, HL (equation 3.17, page 67).

Pearson's χ^2 Statistic

Under the null hypothesis that all aspects of the fitted model are correct, the distribution of P should be

$$P \sim \chi^2_{J-(p+1)} \tag{4.11}$$

where J is the number of distinct covariate patterns and p is the number of parameters in the model. Unfortunately, this does not hold when $J \approx n$ (Hosmer & Lemeshow, 2000). Osius & Rojek (1992) presented a large sample normal approximation for the distribution of P, which involves the following procedure, as detailed in Hosmer & Lemeshow (2000, pg. 153). This allows P to be utilised as a statistic for testing the null hypothesis of adequate model fit. The normal approximation is given for the situation when $J \approx n$.

1. Retain the predicted probabilities (fitted values) from the model, denoted as $\hat{\pi}_j$, $j = 1, 2, \ldots, n$.
2. Create the variable $\nu_j = \hat{\pi}_j(1 - \hat{\pi}_j)$, $j = 1, 2, \ldots, n$.
3. Create the variable $c_j = \frac{(1-2\hat{\pi}_j)}{\nu_j}$, $j = 1, 2, \ldots, n$.
4. Compute the Pearson chi-square statistic shown in equation 3.16, that is,
$$P = \sum_{j=1}^{n} \frac{(y_j - \hat{\pi}_j)^2}{\nu_j}.$$
5. Perform a weighted linear regression of c on $\mathbf{X}$, the model covariates, using the weights ν. Let RSS denote the residual sum-of-squares from this regression, that is,
$$RSS = \sum_{j=1}^{n} (c_j - \hat{c}_j)^2.$$
6. Calculate the standardised statistic
$$z = \frac{[P - (n - p - 1)]}{\sqrt{RSS}}.$$
7. The two-tailed p-value for z can then be calculated using the standard normal distribution, and the null hypothesis (that all aspects of the fitted model are correct) can be formally tested.

Hosmer-Lemeshow Statistic

Under the same null hypothesis assumption of the fitted model being correct, the distribution of the HL statistic is

$$HL \sim \chi^2_{g-2} = \chi^2_8. \tag{4.12}$$

The degrees of freedom $g - 2$ correspond to eight in this case, as the data have been partitioned into $g = 10$ groups, corresponding to decile rankings.

Predictive Ability of the Model

A second way of examining model adequacy is to look at the classification from the fitted model, that is, the predicted outcome of SIDS or no SIDS days in comparison with the observed outcomes. This is dependent on a cut point, c, where if

$$\hat{\pi}_j < c \qquad \hat{y}_j = 0 \text{ (day where no SIDS occurred)}$$
$$\hat{\pi}_j \geqslant c \qquad \hat{y}_j = 1 \text{ (day where at least one SIDS occurred).}$$

As previously, $\hat{\pi}_j$ is the predicted probability (or fitted value). The classification accuracy is quantified by sensitivity (the probability of predicting a day where SIDS occurred given a SIDS did actually occur) and specificity (the probability of predicting no SIDS given that no SIDS actually occurred on that day) (Taube, 1986). The ROC (Receiver Operating Characteristic (Hosmer & Lemeshow, 2000)) curve plots the sensitivity against the compliment of the specificity (1 - specificity) over all possible cut points. The area under the ROC curve also provides a description of the accuracy of the classifications from the fitted model. The area values range from zero to one, and measure the likelihood that a day where at least one SIDS occurred will have a higher value for $P(y = 1))$ than a day where no SIDS occurred. Hosmer & Lemeshow (2000, pg.162) provide a general rule for the discriminative ability of a fitted model for various ROC values:

If ROC $= 0.5$: this suggests no discrimination.
If $0.7 \leqslant$ ROC < 0.8: this is considered acceptable discrimination.
If $0.8 \leqslant$ ROC < 0.9: this is considered excellent discrimination.
If ROC $\geqslant 0.9$: this is considered outstanding discrimination.

4.3.2 Residual Diagnostics

The second stage of assessing the model fit involves examining the contribution to the model of the individual observations looking for influential or poorly fitted values. This is achieved graphically by examining various residual based measures plotted against the predicted probability, $\hat{\pi}_j$.

Pearson Residuals Against the Predicted Probability

An initial plot of the Pearson residuals (equation 3.15, page 66) against $\hat{\pi}_j$ is examined for any obvious discrepancies or unusual patterns.

Influence of the jth Observation

Two further plots, based on statistics that examine the effect of deleting the jth observation on both the model parameters, and the overall goodness-of-fit measures, are utilised to highlight any influential observations.

ΔP against $\hat{\pi}$

Given that the number of distinct covariate patterns $J \approx n$, the Pearson chi-squared statistic that results from the deletion of the $\mathbf{x}_j$th observation is given by

$$\Delta P_j = \frac{r_j^2}{1 - h_j} \tag{4.13}$$

where r_j is the Pearson residual given in equation 3.15 (page 66) and

$$h_j = \hat{\pi}(\mathbf{x}_j)[1 - \hat{\pi}(\mathbf{x}_j)]\mathbf{x}_j'(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_j'. \tag{4.14}$$

Here $\mathbf{x}_j' = [1, x_{1j}, x_{2j}, \ldots, x_{pj}]$ is the vector of covariates corresponding to the jth observation and $\mathbf{V}$ is an n×n diagonal matrix with the jth diagonal $\nu_j = \hat{\pi}(\mathbf{x}_j)[1 - \hat{\pi}(\mathbf{x}_j)]$ (Hosmer & Lemeshow, 2000). In the plot of ΔP against $\hat{\pi}$, observations that have not been well fitted are usually highlighted by points lying in the top corners of the plots.

$\Delta\hat{\beta}$ against $\hat{\pi}$

The change in value of the model parameters, β, when the $\mathbf{x}_j$th observation is deleted is the standardised difference between $\hat{\beta}$ and $\hat{\beta}_{(-j)}$ (Hosmer & Lemeshow, 2000). That is

$$\begin{aligned}\Delta\hat{\beta}_j &= (\hat{\beta} - \hat{\beta}_{(-j)})'(\mathbf{X}'\mathbf{V}\mathbf{X})(\hat{\beta} - \hat{\beta}_{(-j)}) \\ &= \frac{r_j^2 h_j}{(1-h_j)^2}.\end{aligned} \tag{4.15}$$

Outlying values of $\Delta\hat{\beta}$, in the plot of $\Delta\hat{\beta}$ against $\hat{\pi}$, are noted as possible influential observations.

Autocorrelation Function

The final stage of the residual diagnostics comes as a result of the time series nature of the data that is being modelled. Due to this time dependency, serial correlation, otherwise know as autocorrelation, may occur in the residuals, implying that they are not independent. Again, a graphical analysis is utilised, in this case plotting the autocorrelation function (ACF). The ACF is given by

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} \tag{4.16}$$

where $\gamma(\tau)$ =cov$[r(t), r(t+\tau)]$ is the autocovariance coefficient at lag τ. The ACF is a measure of the correlation between residuals $r(t)$ and $r(t+\tau)$. The ACF is then plotted against the lag τ. When visually inspecting the ACF plot, any deviations from a purely random series are highlighted by values outside the range $\pm\frac{2}{\sqrt{n}}$, that is, extreme values (Chatfield, 1996).

4.4 Statistical Modelling Methods

Both the principal component analysis and subsequent PC regression were completed using SAS (*SAS (r) Proprietary Software Release (8.01.01)*), with the FACTOR and LOGISTIC procedures used respectively. The results from the PCA and subsequent modelling are presented separately for Period 1 (1968—1972) and Period 2 (1973—1989). Neither wind direction nor wind chill were included in the PCA, but were incorporated in their original climatic state in the final modelling stages.

The climate variables included in the PCA were those that were found to be significantly related to the incidence of SIDS, over and above seasonality, at an $\alpha \leqslant 0.10$ level of significance. This level was chosen as a way of including all climate variables potentially related to SIDS rates in further analysis, while still reducing the initial dataset of approximately 2500 climatic covariates to a manageable number. A complete list of the climatic variables included in the PCA appears in Table D.1 (Period 1) and Table D.2 (Period 2).

4.5 Results

For improved readability, detailed presentation of the PC sets for both Period 1 and Period 2 have been presented in Appendix C (Tables C.3 - C.9). In these tables, coefficient loadings are presented only for those variables that load highly on a particular PC (where $a_i \geqslant 0.30$). For this reason not all the climate variables

Component number	General Description	Variance Explained	Cumulative Variance
PC1	Wind speed, Dewpoint	0.594	0.594
PC2	Temperature, Wind velocity (North, East, West)	0.033	0.627
PC3	Wind velocity (South, West)	0.031	0.658
PC4	Wind speed	0.030	0.688
PC5	Wind velocity (North, East, West), Pressure	0.026	0.714
PC6	Wind velocity (South), Rain	0.021	0.735
PC7	Wind velocity (North)	0.017	0.752
PC8	Wind velocity (South, West), Pressure, Rain	0.016	0.768
PC9	Wind speed	0.015	0.783
PC10	Wind speed	0.013	0.796
PC11	Wind direction	0.012	0.808
PC12	Wind direction	0.011	0.820
PC13	Wind direction	0.010	0.830

Table 4.1: Details of 13 rotated components (Method 1, Period 1).

appear; some variables have minimal effect on the PC. Similarly some gaps appear in the tables; although a variable may weight highly on PC_i, its loading on PC_j may be negligible.

The cutoff of 0.3 was chosen as it was felt that a loading smaller than 0.3 would mean that the corresponding variable makes only a very small contribution to the PC while large loadings imply an important contribution to the definition of the PC structure. Cutoff values of 0.3 to 0.7 are presented in the literature (for example see Thanasoulias et al. (2003); Thompson et al. (2004); Egan & Angus (2004); Sadeghi (2003)). It is important to note that PC scores (as used in PC regression) were calculated using the loadings from all the variables, and that Tables C.3 to C.9 are presented as an aid to understanding the underlying PC structures.

4.5.1 Period 1 (1968—1972)

The methods used to determine the number of Period 1 climatic PCs to retain and rotate concluded 13, 23 and 16 components were required using Methods 1, 2, and 3 respectively (Section 4.1). Tables 4.1 to 4.3 present a summary of the variables with high component loadings ($a_{ij} \geqslant 0.30$), for each number of rotated components.

The first three components in each set of PCs consist of the same structure, namely wind speed and dewpoint for PC1 (accounting for 59% of the variance explained), temperature, and north, east and west wind velocity for PC2 (3% of the explained variance) and wind velocity from the south and west for PC3 (3% of the explained variance). The fourth component (wind speed) is also common between the 13 and 16 rotated PCs. Other similarities across component structures include the last three PCs from each method being defined by wind direction variables, and the set of PCs from each method containing at least three distinct components defined by wind speed alone. Tables D.3 to D.5 in Appendix D present a detailed summary of the specific variables with high component loadings, and their corresponding loading, for each set of PCs resulting from the three methods of choosing the number of PCs to rotate. The proportion of variance, and cumulative proportion of variance explained for each component is identical across the three sets of PCs.

Wind / Non-wind Partition

A fourth set of components was created, by applying PCA separately to the wind and non-wind based climatic covariates, creating two subsets of PCs. This was done because it was thought, with so many wind based variables, the contribution of other non-wind based variables in the final regression step may be swamped. The criterion described in method 3 was applied to decide how many components to retain, with seven and fourteen for the wind and non-wind based sets respectively. Details of the general structure of these components appear in Table 4.4. Similarly to the previously calculated sets of PCs, wind direction defined the final three PCs in the wind-based PCs, alongside two distinct components defined only by wind

Component number	General Description	Variance Explained	Cumulative Variance
PC1	Wind speed, Dewpoint	0.594	0.594
PC2	Temperature, Wind velocity (North, East)	0.033	0.627
PC3	Wind velocity (South, West)	0.031	0.658
PC4	Temperature, Wind velocity (South, West), Pressure, Rain	0.030	0.688
PC5	Temperature, Wind velocity (North, East, West), Pressure	0.026	0.714
PC6	Wind speed	0.021	0.735
PC7	Rain	0.017	0.752
PC8	Wind speed	0.016	0.768
PC9	Wind velocity (North)	0.015	0.783
PC10	Wind velocity (North)	0.013	0.796
PC11	Wind velocity (South)	0.012	0.808
PC12	Wind velocity (West)	0.011	0.820
PC13	Wind speed	0.010	0.830
PC14	Wind velocity (North)	0.009	0.839
PC15	Temperature, Humidity	0.009	0.848
PC16	Wind velocity (South), Pressure	0.008	0.856
PC17	Wind speed	0.008	0.865
PC18	Pressure	0.008	0.872
PC19	Wind velocity (South), Pressure	0.007	0.879
PC20	Wind speed	0.007	0.887
PC21	Wind direction	0.007	0.893
PC22	Wind direction	0.006	0.899
PC23	Wind direction	0.006	0.905

Table 4.2: Details of 23 rotated components (Method 2, Period 1).

Component number	General Description	Variance Explained	Cumulative Variance
PC1	Wind speed, Dewpoint	0.594	0.594
PC2	Temperature, Wind velocity (North, East)	0.033	0.627
PC3	Wind velocity (South, West)	0.031	0.658
PC4	Wind speed	0.030	0.688
PC5	Temperature, Wind velocity (North, East, West), Pressure	0.026	0.714
PC6	Temperature, Wind velocity (South, West), Rain	0.021	0.735
PC7	Wind velocity (North), Rain	0.017	0.752
PC8	Wind velocity (North)	0.016	0.768
PC9	Wind speed	0.015	0.783
PC10	Wind velocity (South)	0.013	0.796
PC11	Wind speed	0.012	0.808
PC12	Wind velocity (West)	0.011	0.820
PC13	Wind velocity (South), Pressure	0.010	0.830
PC14	Wind direction	0.009	0.839
PC15	Wind direction	0.009	0.848
PC16	Wind direction	0.008	0.856

Table 4.3: Details of 16 rotated components (Method 3, Period 1).

speed. The first two components of the wind-based PCs were comprised of north and west wind velocity and south and west wind velocity for PC1 and PC2 respectively. Dewpoint featured predominantly in the first PC of the non-wind PCs, and along with temperature and rain, defined the first five PCs. Pressure and sunshine individually defined two and four of the final seven PCs respectively. Details of the wind and non-wind PCs, with corresponding component loadings, are presented in Table D.6.

Each set of rotated components was then incorporated into the next stage of the analysis procedure - principal component regression.

Principal Component Regression

Each of the four sets of rotated components found from PCA were modelled separately. Models were of the form

$$\text{logit}[\pi(\mathbf{X}^*)] = \gamma_0 + \gamma_1 \mathbf{Temp_{meanMA30}} + \gamma_2 \mathbf{PC1} + \cdots + \gamma_k \mathbf{PCk}, \tag{4.17}$$

that is, contribution of the PCs to SIDS incidence was examined, over and above the annual cyclic variation, as measured by $Temp_{meanMA30}$.

A backward elimination procedure was implemented to identify components in the regression model that were significantly related to the incidence of SIDS. Initially, all k PCs were included in the model,

Component number	General Description	Cumulative Variance Explained
Wind Components		
PC1	Wind velocity (North, West)	0.681
PC2	Wind velocity (South, West)	0.719
PC3	Wind speed	0.751
PC4	Wind speed	0.782
PC5	Wind direction	0.807
PC6	Wind direction	0.826
PC7	Wind direction	0.842
Non-wind Components		
PC1	Dewpoint	0.216
PC2	Temperature	0.337
PC3	Rain	0.396
PC4	Rain	0.452
PC5	Dewpoint	0.503
PC6	Temperature, Humidity	0.553
PC7	Temperature, Pressure	0.597
PC8	Pressure	0.641
PC9	Pressure	0.677
PC10	Sunshine	0.711
PC11	Pressure	0.737
PC12	Sunshine	0.761
PC13	Pressure	0.785
PC14	Dewpoint	0.807

Table 4.4: Details of wind and non-wind components (Period 1).

	D	AIC	SC	P	HL	Rank sum
Model 1	592.729	608.729	652.680	1763.065	3.589	7
Model 2	592.818	610.818	660.263	1685.631	6.188	13
Model 3	594.056	610.056	654.007	1774.806	11.516	15
Model 4	603.539	617.539	655.996	1764.058	1.319	15

Table 4.5: Goodness of fit statistics for four potential models, Period 1.

alongside $Temp_{meanMA30}$. PCs were removed sequentially, starting with the PC with the largest p-value. This process was continued until all p-values were less than or equal to 0.05. The significant PCs for each set of components are highlighted in Tables 4.1 to 4.4. This process resulted in four competing models for consideration, denoted as follows:

Model 1 - 13 rotated components (derived from Method 1)
Model 2 - 23 rotated components (derived from Method 2)
Model 3 - 16 rotated components (derived from Method 3)
Model 4 - wind and non-wind PC subsets.

The statistics defined in Section 3.2 (page 64) are utilised for comparison between the four models to decide which model best fits the SIDS data. Table 4.5 presents these goodness-of-fit statistics for the four candidate models.

A rank sum method, similar to that applied in Chapter 3, is used to decide which of the four models best fits the SIDS series. This involves ranking the four models by each goodness-of-fit statistic, then summing the ranks for each model. The final rank sums are presented in Table 4.5. Using this procedure, Model 1, with the lowest rank sum of seven, clearly outranks the other three models considered. Parameter estimates, along with corresponding odds ratios for Model 1 are presented in Table 4.6.

Including wind chill (the only Period 1 climate variable not incorporated into the PC structure) into the regression model, did not significantly improve Model 1. It also resulted in the seasonality measure ($Temp_{meanMA30}$) no longer being significant in the model. Therefore, wind chill, is no longer considered at this stage of the analysis of the SIDS—climate relationship in Period 1.

Parameter	Estimate (SE)	Odds ratios (95% CI)	Component description
Intercept	-2.783 (0.370)		
$Temp_{meanMA30}$	-0.054 (0.031)	0.947 (0.891, 1.006)	
PC2	0.327 (0.106)	1.387 (1.127, 1.707)	Temp, North, East, West
PC3	-0.264 (0.130)	0.768 (0.595,0.991)	South, West
PC5	-0.429 (0.149)	0.651 (0.486, 0.873)	North, East, West, Pres
PC6	-0.545 (0.215)	0.580 (0.380,0.885)	South, Rain
PC8	0.218 (0.110)	1.244 (1.003, 1.543)	South, West, Pres, Rain
PC11	-0.281 (0.136)	0.755 (0.579, 0.986)	WindD

Table 4.6: Model 1 parameter details, Period 1.

Assessing the Fit of Model 1

Overall Model Fit

The two statistics utilised to test the overall goodness-of-fit of Model 1 showed no evidence to reject the null hypothesis that all aspects of Model 1 are correct. Specifically, the large sample normal approximation to P, $z = -0.0006$ ($p = 0.999$) and $HL = 3.589$ ($p = 0.892$).

Figure 4.1 presents the sensitivity and specificity for all cut points for Model 1. The low cross-over point between the sensitivity and specificity curves (approximately 0.04) shows that there was good classification for days where no SIDS occurred, and poor classification for days with at least one SIDS death. This is expected with such a low incidence of SIDS: only 77 of the 1797 days in the Period 1 analysis (4.3%) had SIDS occurring.

The ROC curve for Model 1 is presented in Figure 4.2. The area under this curve is 0.707 which corresponds to acceptable discrimination.

These overall goodness-of-fit measures show no evidence that Model 1 inadequately describes the incidence of SIDS in Period 1.

Residual Diagnostics

The Pearson residuals for Model 1 are plotted against the predicted probability, $\hat{\pi}$ in Figure 4.3. Most of the positive residuals correspond to $y_j = 1$, that is, days where at least one SIDS occurred, while conversely, the negative residuals correspond to $y_j = 0$. One point in the top left corner of Figure 4.3 appears to be poorly fitted (highlighted by ◯), with a residual value of 11.41. This corresponds to 13 March 1969, a day where one SIDS death occurred.

Figure 4.4 presents ΔP plotted against the predicted probability, $\hat{\pi}$. As ΔP is essentially the residuals squared, this graph is similar to Figure 4.3, with the poorly fitted point in the top left corner, corresponding to 13 March 1969, further emphasised, with a ΔP value of 130.41 (again, highlighted by ◯).

The diagnostic $\Delta\hat{\beta}$ is plotted against the predicted probability, $\hat{\pi}$, in Figure 4.5. Two values appear to fall slightly away from the main grouping of points, in the top left corner and towards the bottom right corner (again, highlighted by ◯). The first value corresponds to 15 July 1968 while the second point corresponds to 31 October 1972. Both days recorded one SIDS death.

Due to the time series nature of the data, removal of any of the days identified as potential outliers from the dataset is not an option. The covariate patterns corresponding to the three potential outliers are presented in Table 4.7. Each day recorded one SIDS death. This is the only similarity across the three days, and the covariate patterns show no obvious trends that may have resulted in these days being potential outliers.

Figure 4.6 presents the autocorrelation function for Model 1, for lags up to 15 days. This series appears

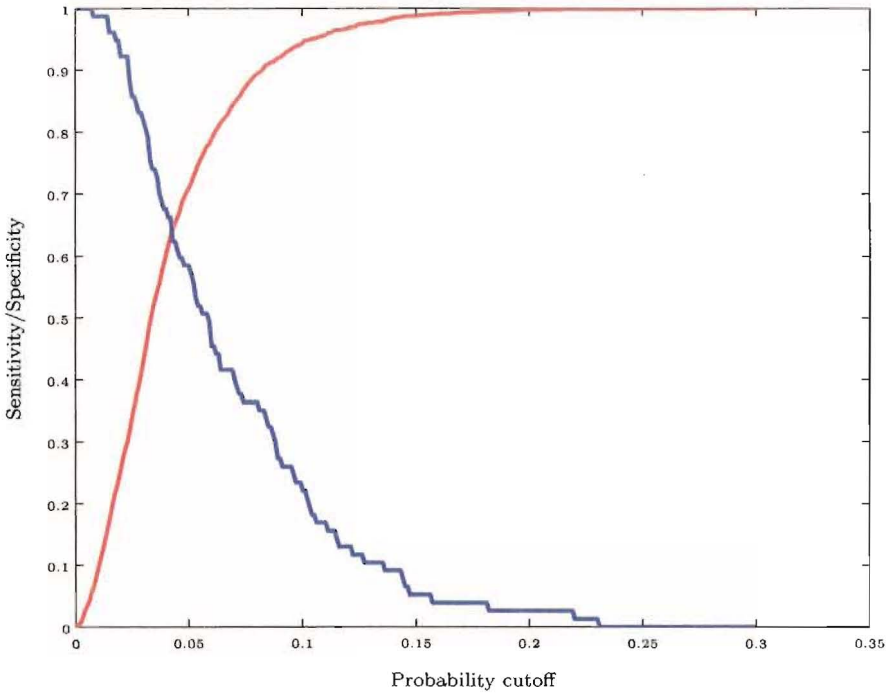


Figure 4.1: Sensitivity and specificity curves for Model 1, Period 1 (— = sensitivity, — = specificity).

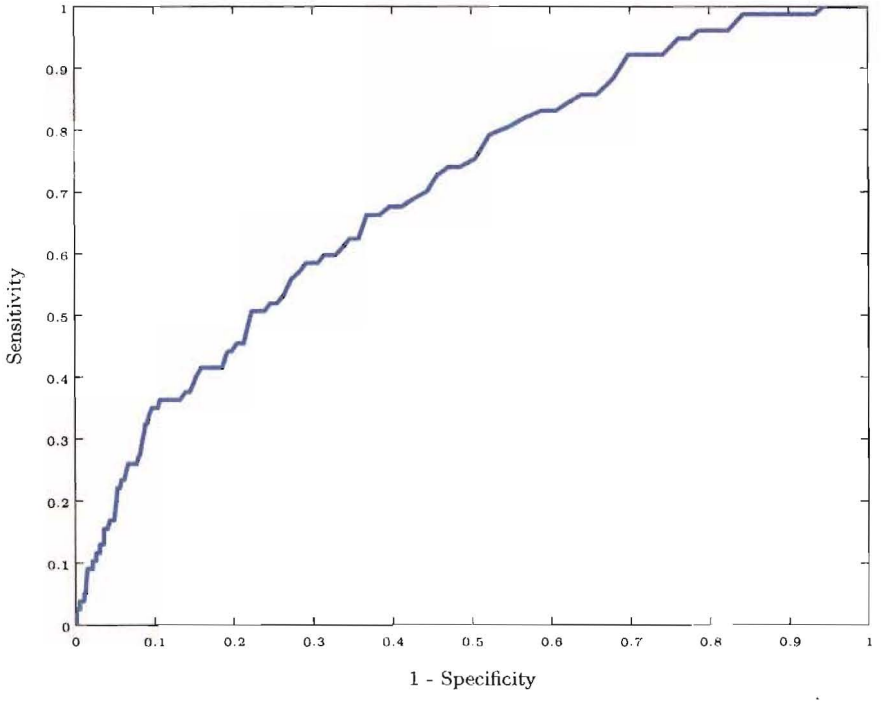


Figure 4.2: ROC curve for Model 1, Period 1.

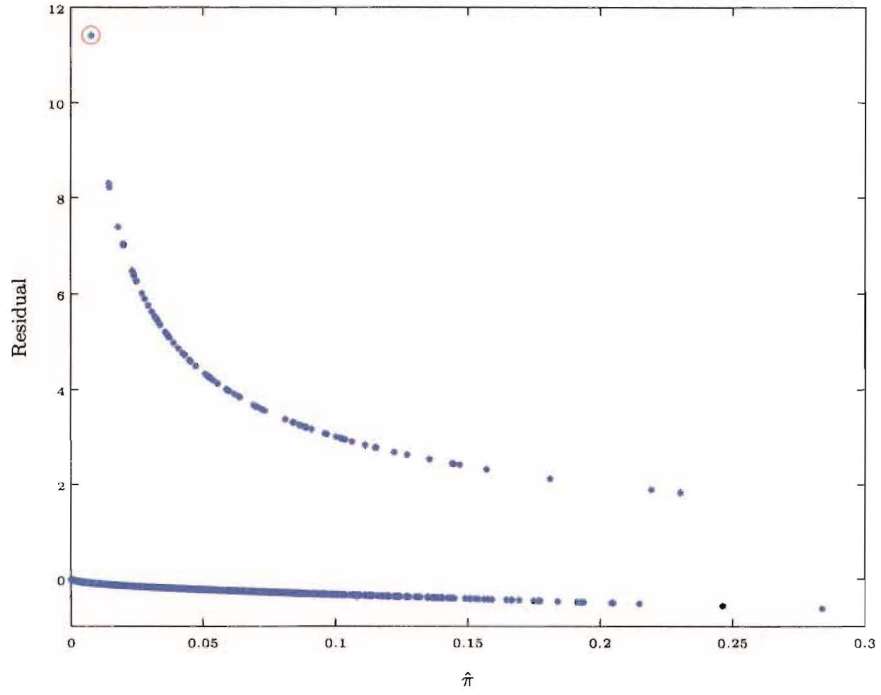


Figure 4.3: Pearson residuals against predicted probabilities for Model 1, Period 1.

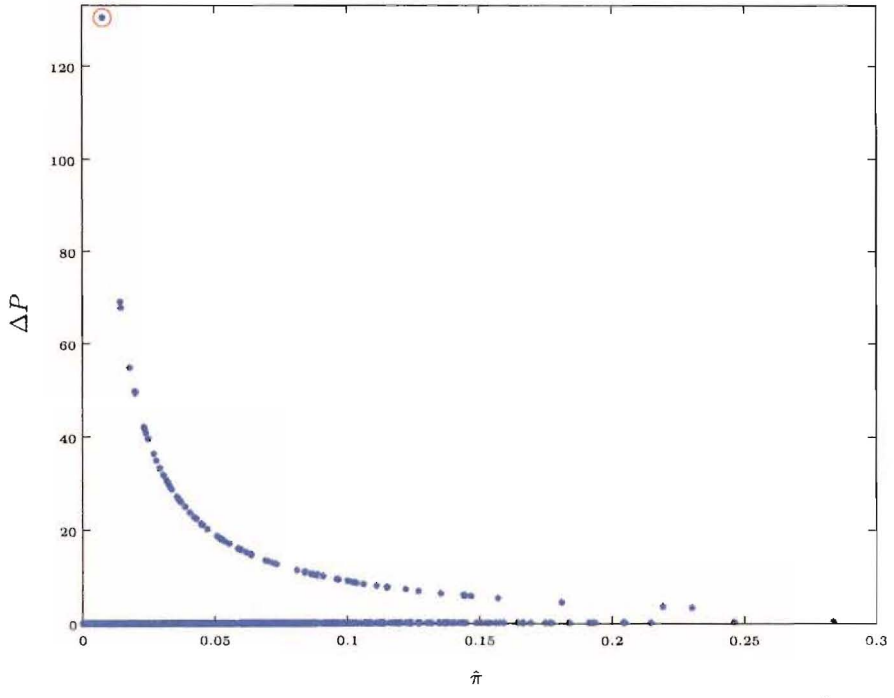


Figure 4.4: ΔP against predicted probabilities for Model 1, Period 1.

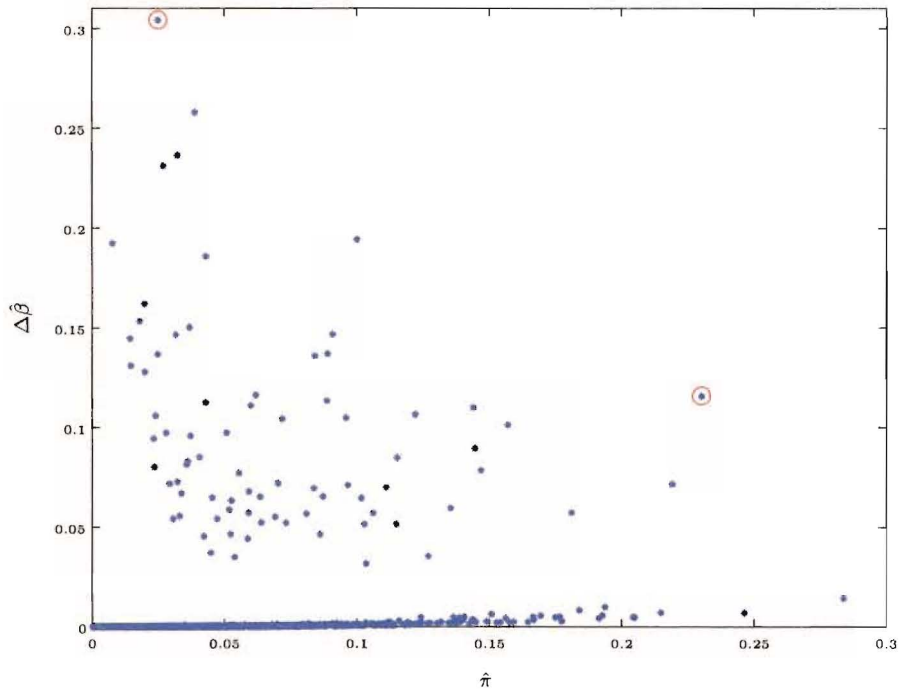


Figure 4.5: $\Delta\hat{\beta}$ against predicted probabilities for Model 1, Period 1.

Date	15 July 1968	13 March 1969	31 October 1972
# SIDS	1	1	1
Season	winter	autumn	spring
$Temp_{meanMA30}$	5.431	15.930	12.005
PC2	0.211	-0.721	3.531
PC3	-0.545	1.740	-1.030
PC5	-1.177	1.523	-0.935
PC6	2.068	-0.622	-0.087
PC8	0.447	-0.466	2.139
PC11	1.002	0.389	0.405

Table 4.7: Covariate values for potentially influential observations, Model 1, Period 1.

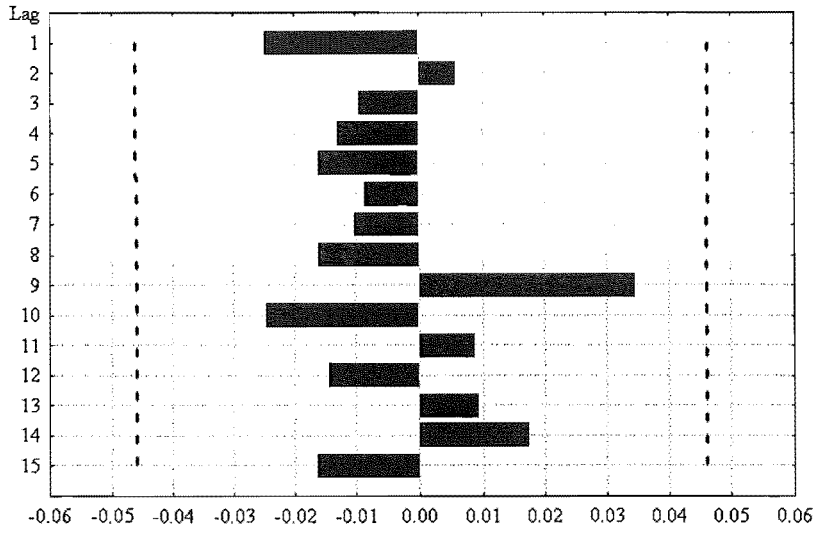


Figure 4.6: Autocorrelation function of the residuals of Model 1, Period 1.

purely random, with no ACF values falling outside the range of $\pm\frac{2}{\sqrt{n}}$. Lags up to one year were also examined, and no obvious deviations from the random series were found (not shown). This implies that serial correlation is not a problem in the residuals from Model 1.

Overall, it appears that Model 1 adequately and appropriately describes the relationship between the incidence of SIDS and climate in Period 1.

Interpretation of Model 1

Interpretation of Model 1 is complicated by the hierarchical-type model created via PC regression: SIDS incidence is predicted by a set of principal components, which in turn, is created from a set of climatic covariates. The final model is pictured diagrammatically in Figure 4.8, to highlight this hierarchy, and aid understanding.

The final (PC) model is given by

$$\begin{aligned}\text{logit}[\hat{\pi}(\mathbf{X}^*)] = &-2.78 - 0.05\mathbf{Temp_{meanMA30}} + 0.33\mathbf{PC2} - 0.26\mathbf{PC3} - 0.43\mathbf{PC5} \\ &- 0.55\mathbf{PC6} + 0.22\mathbf{PC8} - 0.28\mathbf{PC11}. \end{aligned} \tag{4.18}$$

Corresponding standard errors, odds ratios and 95% confidence intervals presented in Table 4.6. The odds ratios infer a decreased risk of SIDS with increasing $Temp_{meanMA30}$, $PC3$ $PC5$, $PC6$, and $PC11$ values. Conversely, SIDS risk increases with increasing values of $PC2$ and $PC8$ (Figure 4.8, Table 4.6).

Figure 4.7 shows the predicted probability of a SIDS occurring by season, for various $Temp_{meanMA30}$ temperatures, holding the other covariates fixed at their seasonal mean values. The seasonal values for $Temp_{meanMA30}$ are highlighted by '$*$'. This illustrates the seasonal risk of SIDS, with deaths more likely to occur in winter. There appears little difference in the mean SIDS risk between the spring, summer, and autumn seasons, as estimated by Model 1.

To interpret the effect of the PCs on SIDS risk, the climate structure of the PCs must be examined. Figure 4.8 presents a diagrammatic illustration of the relationships between the model components and

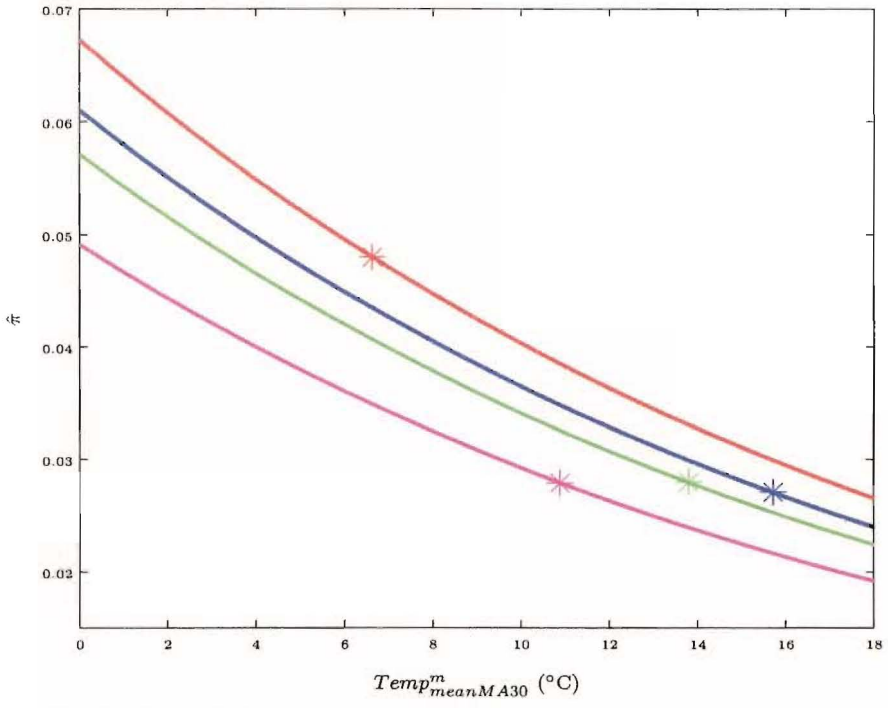


Figure 4.7: Model 1, Period 1, highlighting the effect of $Temp^{m}_{meanMA30}$ on the predicted probability of a SIDS occurring, by season (— Summer, — Autumn, — Winter, — Spring).

corresponding climate structure. All components are essentially a mean of the pictured variables. To record high values of $PC2$ (corresponding to a high risk of SIDS), high values must occur for the $Temp$, $North$, $East$ and $West$ variables that compose the structure of $PC2$. This in turn implies that SIDS risk increases when the temperature on the day of interest increases ($Temp_{minday0}$); when the temperature on $day0$ is higher than it has been over the previous week or fortnight ($Temp_{diffday0-7}$, $Temp_{diffday0-14}$); when the strength and variability in the northern component of wind velocity increases; when the variability in wind velocity from the east increases ($East_{meanAHCday0}$); and when the western wind velocity is stronger and more variable on $day0$ than it was three days previously ($West_{maxday0-3}$, $West_{stdday0-3}$).

A lower risk of SIDS occurs with increasing values of $PC3$. $PC3$, in turn, increases when the south and west wind velocity variables that define the component structure increase. This implies that SIDS risk decreases with increased variability in the south wind velocity two days prior to the day of interest, and when the average wind velocity from the south increased over the previous two days. An increase in the strength and variability of western wind velocity up to two days prior to the day of interest also decreased the risk of SIDS.

Similarly, increasing values of $PC5$ (implying a decreased SIDS risk) relates to an increase in the strength and variability of north, east and west wind velocity, along with pressure seven to eight days prior to $day0$.

An increase in values of $PC6$ (again relating to a lowered risk of SIDS) corresponds to an increase in the strength and variability of wind velocity from the south seven days prior to $day0$. Also, an increase in the intensity and variability of rainfall up to a week before the day of interest increases $PC6$.

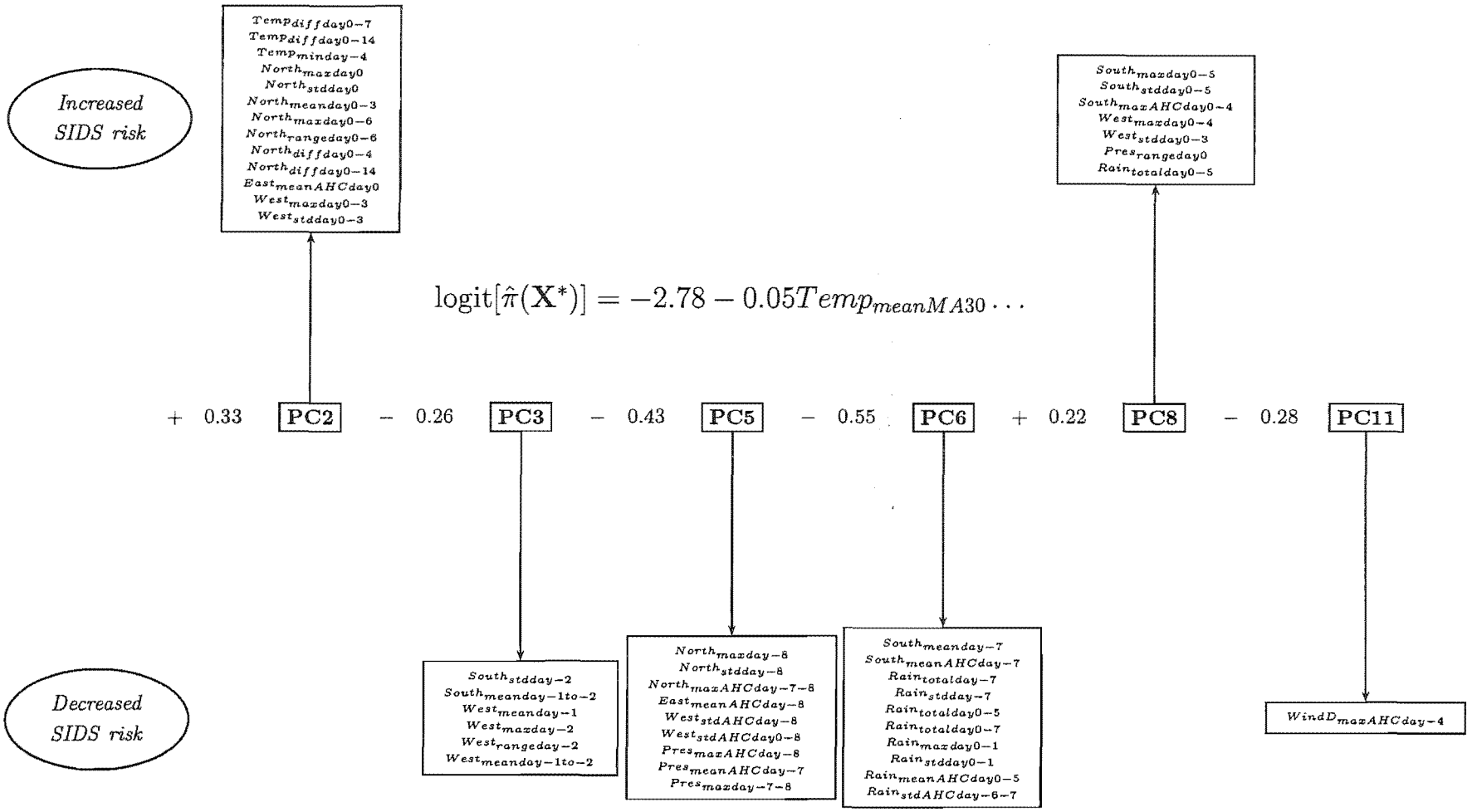


Figure 4.8: A pictorial illustration of the effects of increasing a model component (holding other components fixed) on the resulting risk of SIDS (Model 1, Period 1, 'd').

Component number	General Description	Variance Explained	Cumulative Variance
PC1	Temperature, Wind velocity (South), Rain, Sunshine, Radiation	0.310	0.310
PC2	Wind velocity (South, East, West), Pressure	0.209	0.519
PC3	Wind velocity (North), Humidity, Pressure	0.057	0.576
PC4	Wind velocity (South)	0.044	0.620
PC5	Temperature, Wind velocity (North, West), Humidity, Wind direction	0.037	0.657
PC6	Temperature, Humidity	0.031	0.688
PC7	Wind velocity (West), Humidity, Pressure, Dewpoint	0.023	0.711
PC8	Wind velocity (West), Humidity, Radiation	0.019	0.730
PC9	Humidity	0.017	0.747
PC10	Temperature, Humidity	0.014	0.761
PC11	Wind direction	0.013	0.774
PC12	Wind direction	0.013	0.787
PC13	Wind direction	0.012	0.799
PC14	Humidity	0.011	0.810
PC15	Wind direction	0.010	0.820

Table 4.8: Details of 15 rotated components (Method 1, Period 2).

$PC8$, corresponds to increased risk of SIDS with increasing values. This occurs when the difference between $day0$ and four or five days previously increases for south and west wind velocity strength and variability, and also the total rainfall. An increase in the range of pressure on $day0$ also causes an increase in $PC8$.

The structure of $PC11$ comprises solely of $WindD_{maxAHCday-4}$. Therefore an increase in the variability in wind direction four days prior to $day0$ relates to an increase in PC11, which in turn corresponds to a decreased risk of SIDS.

4.5.2 Period 2 (1973—1989)

The three methods described in Section 4.1 to determine the number of PCs to retain and rotate returned 15, 26 and 25 components for method 1, method 2 and method 3 respectively. Tables 4.8 to 4.10 present a general overview of the component structure corresponding to those climatic variables with high loadings for each number of rotated components.

As with the PCs created for Period 1, there are many similarities among the component structures of each set of PCs for Period 2. The first two components in each set consist of an identical structure, namely temperature, southern wind velocity, rain, sunshine and solar radiation for PC1 (accounting for 31% of the explained variance) and south, east and west wind velocity with pressure for PC2 (21% of the explained variance). Other similarities across all three PC sets include at least three components structurally consisting solely of wind direction, and a further two components composed solely of humidity, while temperature was combined with humidity in two components. The two PC sets consisting of 25 and 26 components were very similar in structure. The proportion of variance, and cumulative proportion of variance explained for each component is identical across the three sets of PCs.

Wind / Non-wind Partition

A fourth set of components was again created by partitioning the wind and non-wind based variables into two subsets of PCs. A general description of the resulting sets is presented in Table 4.11. This table shows that 7 and 19 components were selected to be retained and rotated for the wind and non-wind based sets respectively. As with the previously calculated sets of PCs, wind direction defined four of the seven wind based components, with the remaining three having a structure comprised of various wind velocity variables. Humidity, pressure and temperature made up the structure of the final eleven PCs in the non-wind based set, either individually or in combination.

Tables D.7 to D.10 in Appendix D present component loading details for the four PC sets for the specific variables that define the PC structures.

Component number	General Description	Variance Explained	Cumulative Variance
PC1	Temperature, Wind velocity (South), Rain, Sunshine, Radiation	0.310	0.310
PC2	Wind velocity (South, East, West), Pressure	0.209	0.519
PC3	Wind speed, Dewpoint	0.057	0.576
PC4	Wind velocity (North, East, West), Pressure	0.044	0.620
PC5	Pressure	0.037	0.657
PC6	Wind velocity (South, West), Pressure, Sunshine	0.031	0.688
PC7	Temperature, Wind velocity (North, West), Pressure	0.023	0.711
PC8	Wind speed	0.019	0.730
PC9	Temperature, Wind velocity (West) Pressure, Sunshine, Radiation	0.017	0.747
PC10	Temperature, Humidity	0.014	0.761
PC11	Wind velocity (North), Pressure	0.013	0.774
PC12	Wind velocity (South)	0.013	0.787
PC13	Humidity	0.012	0.799
PC14	Wind velocity (West), Humidity	0.011	0.810
PC15	Wind velocity (North), Humidity	0.010	0.820
PC16	Temperature, Humidity	0.010	0.829
PC17	Wind velocity (North), Humidity	0.009	0.839
PC18	Temperature, Humidity	0.009	0.847
PC19	Wind direction	0.008	0.856
PC20	Temperature, Wind direction	0.008	0.864
PC21	Humidity	0.007	0.871
PC22	Pressure	0.007	0.878
PC23	Wind direction	0.007	0.885
PC24	Wind direction	0.006	0.892
PC25	Wind direction	0.006	0.898
PC26	Humidity	0.006	0.904

Table 4.9: Details of 26 rotated components (Method 2, Period 2).

Component number	General Description	Variance Explained	Cumulative Variance
PC1	Temperature, Wind velocity (South), Rain, Sunshine, Radiation	0.310	0.310
PC2	Wind velocity (South, East, West), Pressure	0.209	0.519
PC3	Wind speed, Dewpoint	0.057	0.576
PC4	Wind velocity (North, East, West), Pressure	0.044	0.620
PC5	Pressure	0.037	0.657
PC6	Wind velocity (South, West), Pressure, Sunshine	0.031	0.688
PC7	Temperature, Wind velocity (North, West), Pressure	0.023	0.711
PC8	Wind speed	0.019	0.730
PC9	Temperature, Sunshine, Radiation	0.017	0.747
PC10	Wind velocity (North), Humidity, Pressure	0.014	0.761
PC11	Temperature, Humidity	0.013	0.774
PC12	Wind velocity (South)	0.013	0.787
PC13	Humidity	0.012	0.799
PC14	Wind velocity (West), Humidity	0.011	0.810
PC15	Wind velocity (North), Humidity	0.010	0.820
PC16	Wind velocity (North), Humidity	0.010	0.829
PC17	Temperature, Humidity	0.009	0.839
PC18	Temperature, Humidity	0.009	0.847
PC19	Wind direction	0.008	0.856
PC20	Temperature, Wind direction	0.008	0.864
PC21	Pressure	0.007	0.871
PC22	Wind direction	0.007	0.878
PC23	Wind direction	0.007	0.885
PC24	Wind direction	0.006	0.892
PC25	Humidity	0.006	0.89

Table 4.10: Details of 25 rotated components (Method 3, Period 2).

Component number	General Description	Cumulative Variance Explained
Wind Components		
PC1	Wind velocity (South, East, West)	0.456
PC2	Wind velocity (South)	0.766
PC3	Wind direction	0.803
PC4	Wind direction	0.825
PC5	Wind velocity (North), Wind direction	0.845
PC6	Wind direction	0.862
PC7	Wind direction	0.877
Non-wind Components		
PC1	Temperature, Pressure, Rain, Sunshine	0.172
PC2	Pressure, Sunshine, Radiation	0.296
PC3	Dewpoint	0.404
PC4	Pressure	0.495
PC5	Temperature, Sunshine, Radiation	0.551
PC6	Temperature, Humidity	0.597
PC7	Sunshine, Radiation	0.638
PC8	Pressure	0.670
PC9	Temperature, Humidity	0.700
PC10	Temperature, Humidity	0.727
PC11	Humidity	0.754
PC12	Humidity	0.780
PC13	Temperature	0.804
PC14	Humidity	0.825
PC15	Humidity	0.845
PC16	Humidity	0.863
PC17	Pressure	0.880
PC18	Pressure	0.892
PC19	Humidity	0.903

Table 4.11: Details of wind and non-wind based components, Period 2.

Structural Differences Between Period 1 and Period 2

Comparing the structure of the climatic components created for Period 2 with the Period 1 based components showed some overall differences between time periods. For example, the first PC in Period 1, comprising approximately 60% of the variance explained by the PCs, consisted of wind speed and dewpoint, whereas in Period 2, the first PC was an amalgamation of temperature, south wind velocity, rain, sunshine and radiation, and only accounted for 31% of the explained variance. The wind speed-dewpoint component was the third PC in the 25 and 26 component sets in Period 2. Sunshine and radiation appear more often in the component structures of the PCs from Period 2 than Period 1, along with PC structures defined solely by humidity.

Principal Component Regression

As with Period 1, each of the four sets of rotated PCs found from the PCA were modelled separately. Models were of the following form:

$$\text{logit}[\pi(\mathbf{X}^*)] = \gamma_0 + \gamma_1 \mathbf{sin}(\frac{\mathbf{2\pi t}}{\mathbf{365}}) + \gamma_2 \mathbf{cos}(\frac{\mathbf{2\pi t}}{\mathbf{365}}) + \gamma_3 \mathbf{PC1} + \cdots + \gamma_k \mathbf{PCk}. \tag{4.19}$$

In this period, the seasonal pattern of SIDS incidence was modelled by the combination of sine and cosine functions. The contribution of the PCs to the estimated probability of SIDS occurrence was examined, over and above the annual cyclic variation. A backwards elimination procedure was again implemented, as described previously for Period 1. Those PCs found to be significantly related to the incidence of SIDS ($\alpha \leqslant 0.05$) are highlighted in Tables 4.8 to 4.11. The resulting four competing models are denoted as follows:

Model 1 - 15 rotated components (derived from Method 1)

Model 2 - 26 rotated components (derived from Method 2)

Model 3 - 25 rotated components (derived from Method 3)

Model 4 - wind and non-wind PC subsets.

Table 4.12 presents the goodness-of-fit statistics, defined in Section 3.2 (page 64), alongside the final rank sum for each of the competing models. Model 2, with the lowest final rank sum of ten, is taken as the

	D	AIC	SC	P	HL	Rank sum
Model 1	3207.237	3229.237	3296.575	6108.730	6.224	13
Model 2	3188.572	3214.572	3302.111	6172.032	2.486	10
Model 3	3182.610	3210.610	3304.883	6195.255	5.236	12
Model 4	3190.102	3216.102	3303.641	6133.514	15.489	15

Table 4.12: Goodness of fit statistics for four potential models, Period 2.

Parameter	Estimate (SE)	Odds ratios (95% CI)	Component description
Intercept	-2.641 (0.054)		
$\sin(\frac{2\pi t}{365})$	−0.223 (0.071)	0.800 (0.697, 0.918)	
$\cos(\frac{2\pi t}{365})$	−0.510 (0.073)	0.600 (0.521, 0.692)	
PC2	−0.136 (0.050)	0.873 (0.791,0.963)	South, East, West, Pres
PC3	−0.116 (0.049)	0.891 (0.809, 0.980)	WindS, Dew
PC8	−0.153 (0.053)	0.859 (0.774,0.952)	WindS
PC9	0.155 (0.053)	1.167 (1.053, 1.294)	Temp, West, Pres, Sun, Rad
PC10	−0.114 (0.049)	0.893 (0.811, 0.983)	Temp, Humid
PC13	−0.154 (0.058)	0.858 (0.766, 0.961)	Humid
PC16	−0.106 (0.052)	0.899 (0.821, 0.982)	Temp, Humid
PC17	−0.107 (0.046)	0.898 (0.821, 0.982)	North, Humid
PC19	−0.109 (0.051)	0.897 (0.812, 0.991)	WindD
PC22	−0.105 (0.050)	0.901 (0.817, 0.993)	Pres

Table 4.13: Model 2 parameter details, Period 2.

preferred model, over the other models with higher final rank sums. Parameter estimates, along with odds ratios and a brief component description are presented in Table 4.13 for Model 2.

Addition of those climate variables not included in the PCs into the PC regression model failed to identify any further significant associations with SIDS, with the exception of $WindD_{day-7}$. The calm category (coded 1) was found to be significantly related to the incidence of SIDS. After comparison of the goodness-of-fit statistics from this model against Model 2, Model 2 remained the preferred choice. The values for the model which included $WindD_{day-7}$ were: $D = 3190.102$, $AIC = 3221.903$, $SC = 3370.046$ and $HL = 6.3817$. Model 2 is therefore examined in more detail in terms of adequacy and fit with respect to estimating the probability of SIDS.

Assessing the Fit of Model 2

Overall Model Fit

The large sample normal approximation to P, $z = -0.017$ ($p = 0.798$) and the Hosmer-Lemeshow goodness-of-fit statistic, $HL = 2.486$ ($p = 0.962$) both showed no evidence to reject the null hypothesis that the fitted model, Model 2, is correct in all aspects.

The sensitivity and specificity values for Model 2, over all cut points from zero to one, is pictured in Figure 4.9. In comparison with Figure 4.1 (the sensitivity and specificity curves for Model 1, Period 1), this model shows a marginal improvement, with the point of cross-over between the two curves being at a cut point of approximately 0.075 as compared to approximately 0.04 for Period 1. As previously, the classification for days where no SIDS occurred is good (specificity) while classification for days with at least

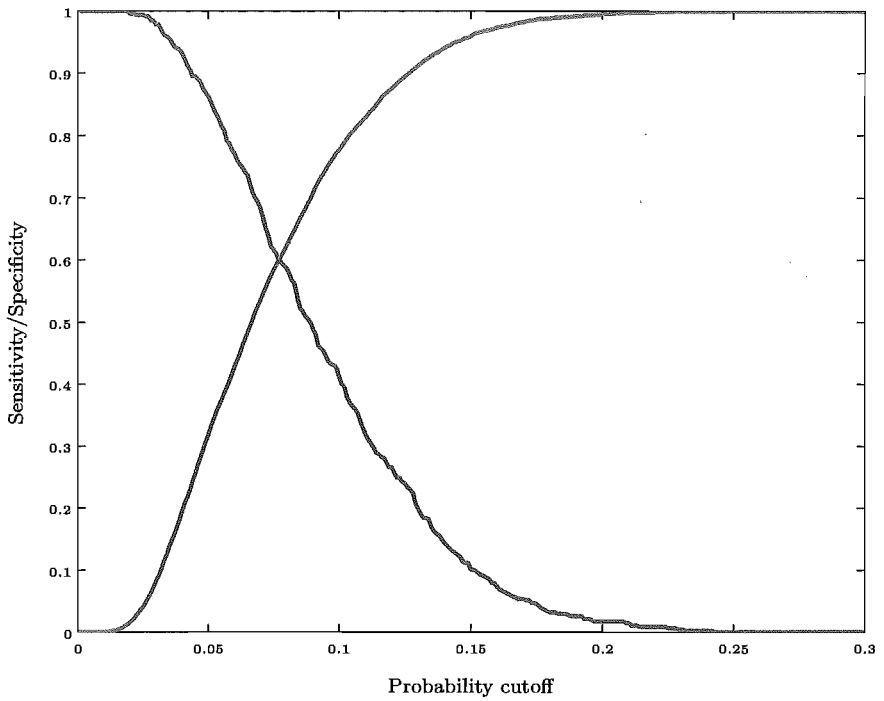


Figure 4.9: Sensitivity and specificity curves for Model 2, Period 2 (— = sensitivity, ---- = specificity).

SIDS death is poor (sensitivity), indicated by the low cross-over value. The SIDS incidence was again very low in this period, with only 467 days where a death occurred out of the 6209 days in Period 2 (7.5%).

Figure 4.10 presents the ROC curve for Model 2. The area under the curve, which quantifies the classification ability of the fitted model, is 0.65. This falls below the acceptable discrimination bracket (0.7 to 0.8).

Overall, these measures of goodness-of-fit show no evidence against Model 2 adequately fitting the incidence of SIDS in Period 2.

Residual Diagnostics

Residual values are examined, to identify any potentially influencing observations. Figure 4.11 presents the Pearson residuals plotted against the predicted probability $\hat{\pi}$. As with Figure 4.3 (Pearson residuals for Model 1, Period 1), there is an obvious split between the values corresponding to a day with no SIDS death (the negative residuals) and those corresponding to days with at least one SIDS death (the positive residuals). No observations appear to be separate from the general trend.

Figure 4.12 shows ΔP plotted against the predicted probability $\hat{\pi}$. This highlights the same trends as seen in Figure 4.11, with no obvious outliers.

In the plot of $\hat{\beta}$ against $\hat{\pi}$, one observation appears to fall away from the main grouping. With a $\hat{\beta}_j$ value of 0.078, removal of this value, corresponding to 30 May 1974, is not considered to unduly effect the model parameters.

The autocorrelation function for Model 2 is presented in Figure 4.14. As with Model 1 in Period 1 (Figure 4.6), this does not show any evidence of a non-random pattern in the time series of residuals.

Overall, it appears that Model 2 adequately and appropriately describes the relationship between climate and the incidence of SIDS in Period 2.

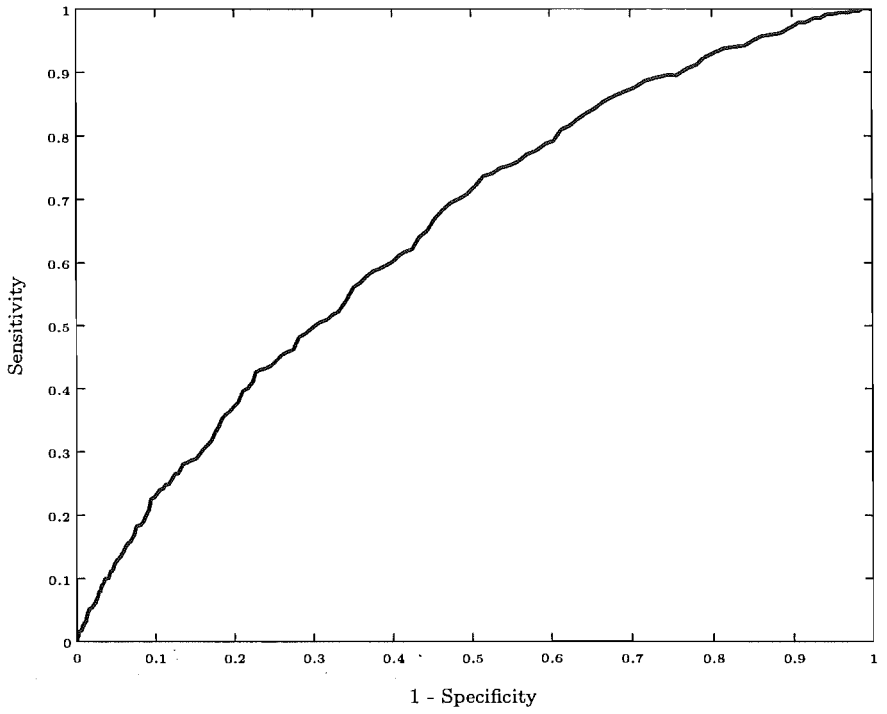


Figure 4.10: ROC curve for Model 2, Period 2.

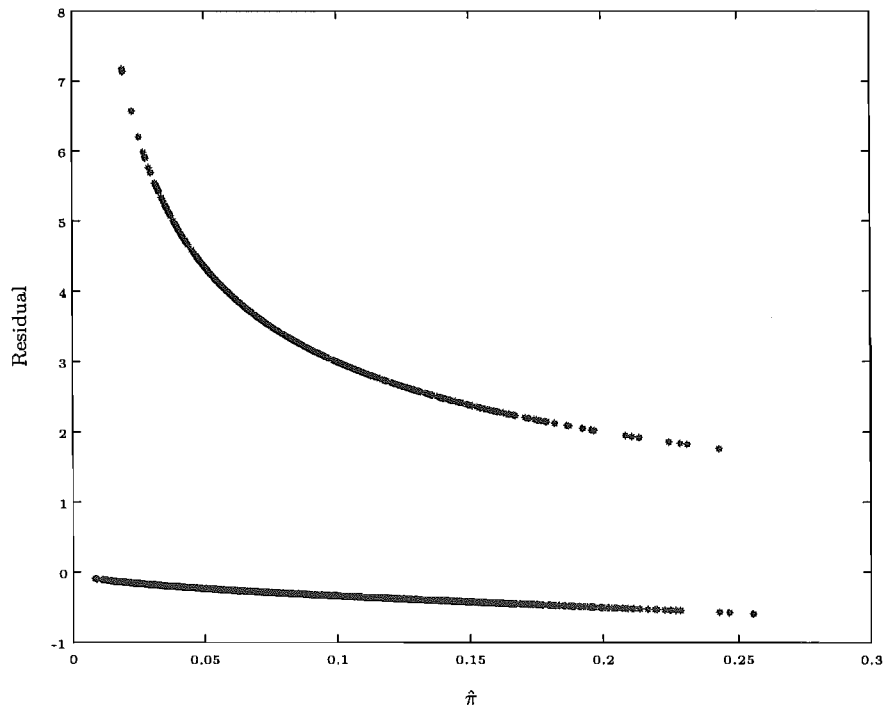


Figure 4.11: Pearson residuals against predicted probabilities for Model 2, Period 2.

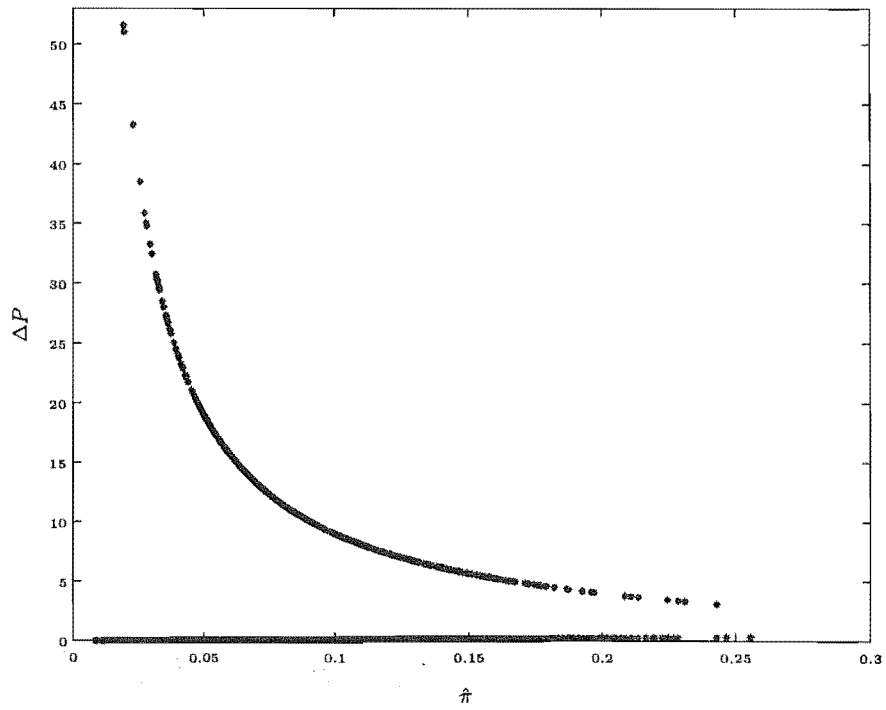


Figure 4.12: ΔP against predicted probabilities for Model 2, Period 2.

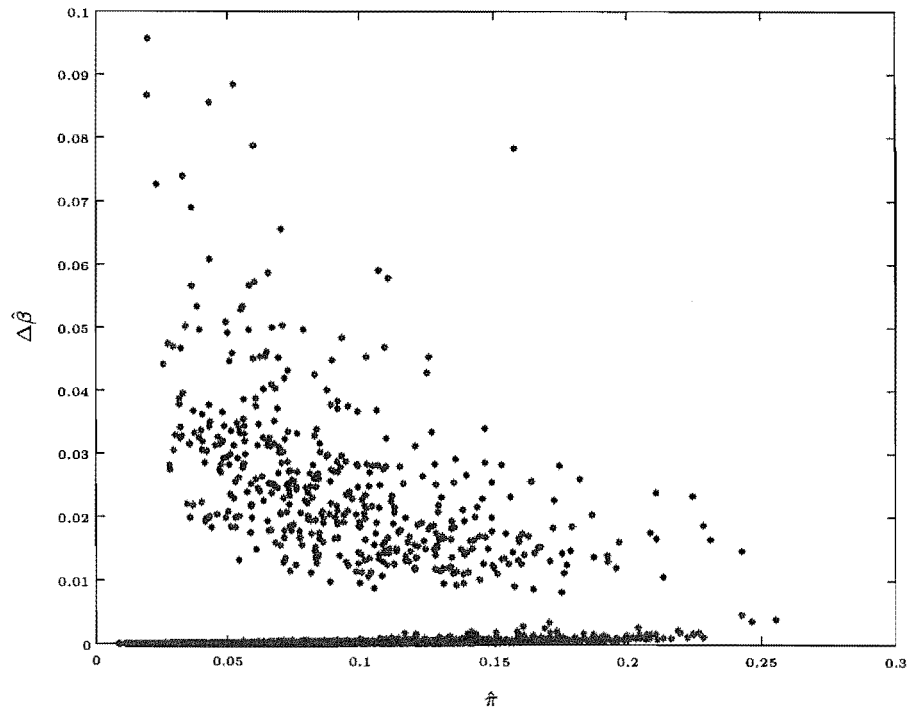


Figure 4.13: $\Delta\hat{\beta}$ against predicted values for Model 2, Period 2.

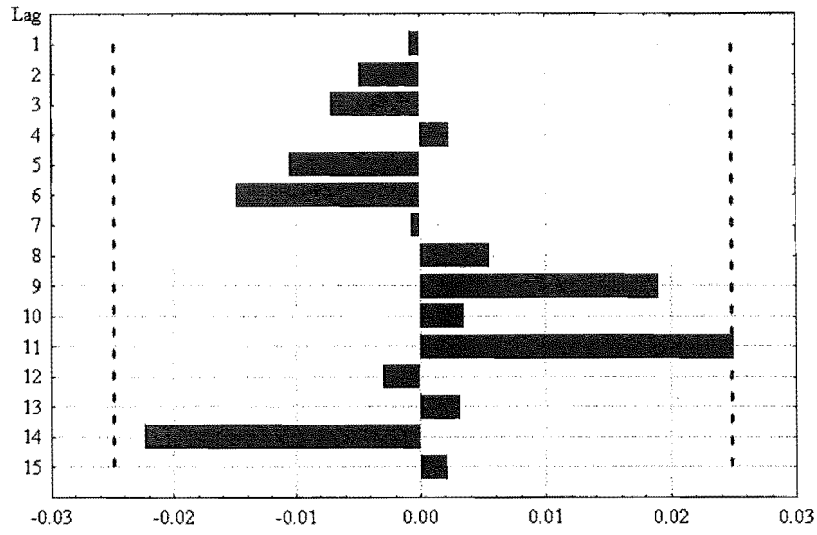


Figure 4.14: Autocorrelation function of the residuals of Model 2, Period 2.

Interpretation of Model 2

The final model chosen to describe the daily relationship between climate and the incidence of SIDS in Period 2 is

$$\begin{aligned}\text{logit}[\hat{\pi}(\mathbf{X}^*)] = &-2.64 - 0.22\sin(\frac{2\pi \mathbf{t}}{365}) + 0.51\cos(\frac{2\pi \mathbf{t}}{365}) - 0.14\mathbf{PC2} - 0.12\mathbf{PC3} - 0.15\mathbf{PC8} \\ &+ 0.16\mathbf{PC9} - 0.11\mathbf{PC10} - 0.15\mathbf{PC13} - 0.11\mathbf{PC16} - 0.11\mathbf{PC17} \\ &- 0.11\mathbf{PC19} - 0.11\mathbf{PC22}, \end{aligned} \tag{4.20}$$

which is shown diagrammatically in Figure 4.15. This figure presents the hierarchy involved in this complex model. The structure of the components range from a single climatic variable to a weighted mean of fourteen climate measures. Hence, to interpret this model in any practical way is difficult. Therefore, interpretation is restricted to an overview of the relationships between the incidence of SIDS, the model components and the climatic structure of those components.

The combination of the sine and cosine curves results in a sinusoid function which describes the seasonal fluctuations in the incidence of SIDS. Of the ten components in the model, $PC9$ is the only component that, when increased, results in an increase in the risk of SIDS. $PC9$, in turn, increases when the temperature, western wind velocity, pressure, sunshine and radiation variables that define its structure increase. The remaining components related to a decreased risk of SIDS with decreasing values of the PCs.

$PC2$ is essentially defined by various wind velocity variables; $PC3$ is structurally defined by dewpoint; $PC8$ is explicitly defined by varying measures of wind speed; and $PC13$ is defined by two humidity variables. Increasing the values of $PC2$, $PC3$, $PC8$, $PC10$, $PC13$, $PC16$, $PC17$, $PC19$ or $PC22$ corresponds to a decrease in the risk of SIDS. For instance, increased values of absolute hourly change in humidity over the five days prior to $day0$ ($PC13$), increased variability in wind direction three days before the day of interest ($PC19$), or an increase in the pressure difference between $day0$ and $day - 8$ ($PC22$) would each infer an increased SIDS risk. These relationships are summarised in Figure 4.15.

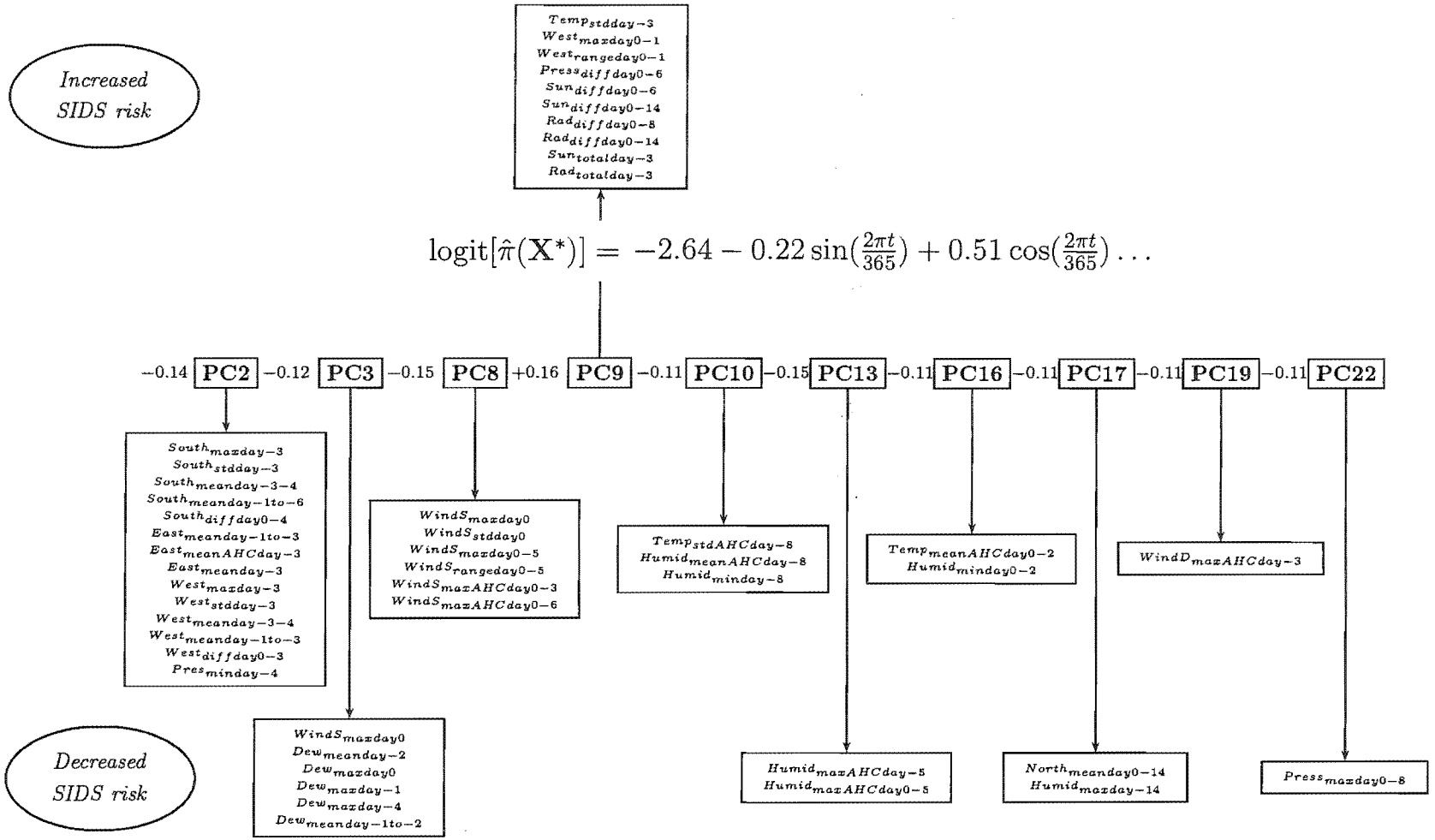


Figure 4.15: A pictorial illustration of the effects of increasing a model component (holding other components fixed) on the resulting risk of SIDS (Model 2, Period 2).

4.6 Summary and Conclusion

This chapter has resulted in models which describe the incidence of SIDS with respect to climate. An overview of these models was presented in Section 4.5, yet creating a climatic summary of what constitutes a 'high-risk' SIDS day, or on the other hand, a 'low-risk' SIDS day was not possible.

The PC regression models were created in a way which removes the problem of strong interrelationships which are inherent within varying measures of the weather. This was achieved by first creating principal components of the climate measures, which, by the nature of their construction, are uncorrelated. These components were then incorporated, via principal component regression, into the multiple logistic model. The two resulting models (for Period 1 and Period 2) were examined in terms of their appropriateness and ability to describe the incidence of SIDS, alongside residual analyses aimed at identifying any potentially influential observations or serial correlations. No evidence was found against either model, suggesting that they were both appropriate and adequate to describe the relationship between the incidence of SIDS and climate patterns.

The disadvantage of this technique is the difficulty making practical and meaningful interpretations from the hierarchical structure underlying the component built model. Though the profile of SIDS risk was adequately captured through the PC based covariates incorporated in the PC regression models, the added dimension involved in the underlying climatic structure of the PCs made interpreting these models a complex issue. This is a similar issue to that of the logistic regression model containing multiple predictors (discussed in Chapter 3), where such a large number of climate-based regressors were contained in the models (25 and 35 for Period 1 and Period 2 respectively), that interpreting the models in terms of the risk profile of SIDS with respect to climate, was not practically possible. Therefore, the next stage of analysis involves modelling the incidence of SIDS at a monthly level. Information loss will necessarily be involved, but will be balanced against the prospect of creating more interpretable monthly climatic covariates. It is hoped that this will lead to a more understandable and replicable model which describes the relationship between SIDS and climate.

Chapter 5

Conversion of Daily Variables to a Monthly Level

This chapter describes the conversion methods used, and variables created, to summarise the daily SIDS and climate data onto a monthly time frame. In the previous chapters (Chapters 3 and 4), a comprehensive analysis of a multitude of daily climatic variables were examined in relation to the incidence of SIDS. This lead to models that were statistically sound, but difficult to interpret in a way that was sensible and meaningful. This chapter reduces the available information to a monthly level, with the primary aim of working with variables that are intuitively useful and contain interpretable information. Monthly climate variables are thus defined in a way which allows a practical interpretation.

When modelling the daily data, daily, within day and between day climate changes, over and above the gross seasonal fluctuations were able to be detected. Converting to the monthly level is a process which involves balancing the information lost in the summarisation process against the practical gains of relating statistical models back to meaningful trends. The resulting monthly models will therefore allow an interpretable climatic profile of SIDS risk to be identified.

SIDS

The monthly SIDS variable, denoted $SIDS_{month}(i)$ or $y^m(i)$, where y represents the response variable, is defined as the number of SIDS deaths in calendar month i. Note $i = 1, \ldots, 59$ for Period 1, representing February 1968 through to December 1972, and for Period 2, $i = 1, \ldots 204$, corresponding to January 1973 through to December 1989. Figure 5.1 shows the monthly SIDS series, over the two periods. Numbers peak over the years 1980 to 1988, with the nine SIDS in one month, the maximum value, recorded in June 1983. The minimum number of SIDS in one month was zero, recorded over 43 months, or 16% of the months in Periods 1 and 2. The mean number of SIDS per month is 2.15, with corresponding variance of 2.89 over the 22 year time frame. By period, this equates to the following descriptive statistic: *Period 1* mean = 1.33, variance = 1.24; *Period 2* mean = 2.39, variance = 3.14.

Population at Risk (NAR)

It is widely known that there are annual trends in birth numbers (see Figure 5.2), and also changes over time (Lam & Miron, 1994; Grech et al., 2003). Both these trends will have an effect on the incidence of SIDS, as a result of more, or less, infants in the postnatal age group (one to twelve months). The following variable was created to account for these underlying population trends.

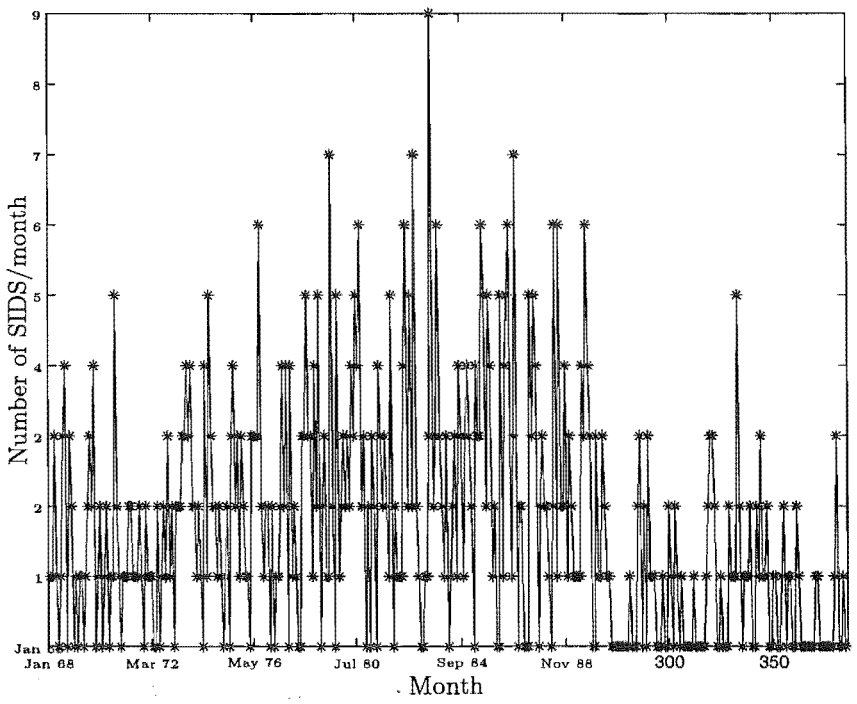


Figure 5.1: Number of SIDS per month, 1968—1989.

For month i the population at risk of SIDS is defined as all infants who are aged between 1 and 12 months (the postnatal age period) in month i. The number of infants at risk, denoted NAR, was estimated. This was necessary as information on birth numbers in Canterbury at a monthly level was unavailable. Annual Canterbury birth numbers were available for the period of the study, alongside monthly details for the thirteen year period commencing from 1980.

From the available monthly birth data, the distribution of births over a year was estimated by calculating the median number of births for each month over the 13 years of data. This estimated birth distribution is shown in Figure 5.2, which shows a peak in birth numbers occurring in September. August and October also have a high birth rate.

The estimated monthly distribution of birth numbers was then applied to those years where the monthly data was not available. This was achieved by simply splitting the annual birth numbers into the estimated monthly percentages. The full monthly birth series is given in Figure 5.3 for Period 1 and 2. This shows a decrease in the number of births annually from 1973 to 1981, with numbers appearing reasonably steady from 1981 onwards, with the exception of a pronounced peak in 1988.

NAR

Finally, to calculate $NAR(i)$, the number of infants born two to twelve months previously were summed, as these infants are aged approximately between one and twelve months at month i. That is

$$NAR(i) = \sum_{j=i-12}^{i-2} \text{birth}(j) \tag{5.1}$$

where birth(i) represents the number of infants born in month i. For example, the number of infants at risk of SIDS in January 1973 are all those infants that were born between January 1972 and November 1972, as these infants will be in the postnatal age period.

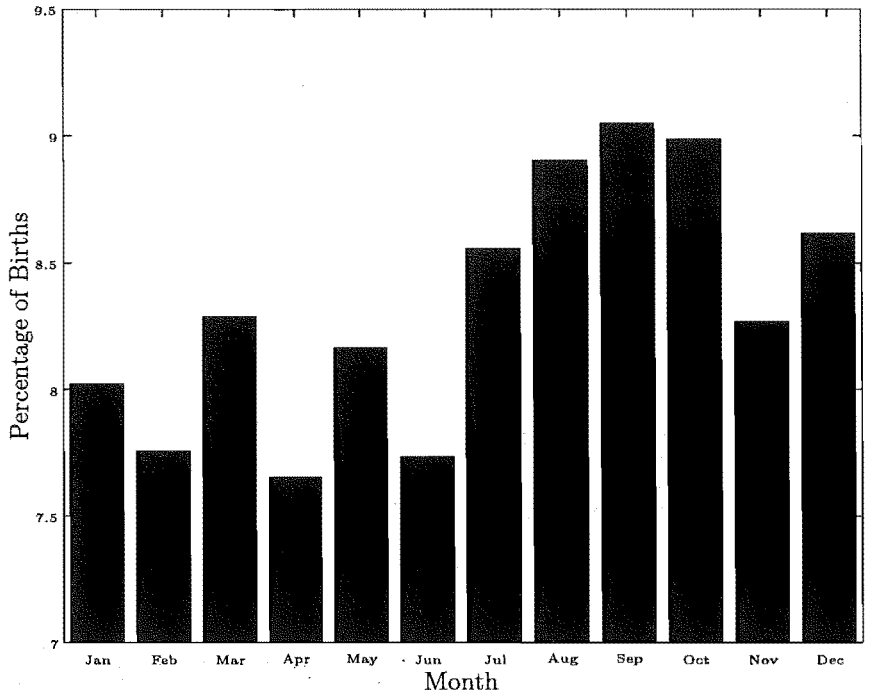


Figure 5.2: Estimated monthly distribution of Canterbury birth numbers.

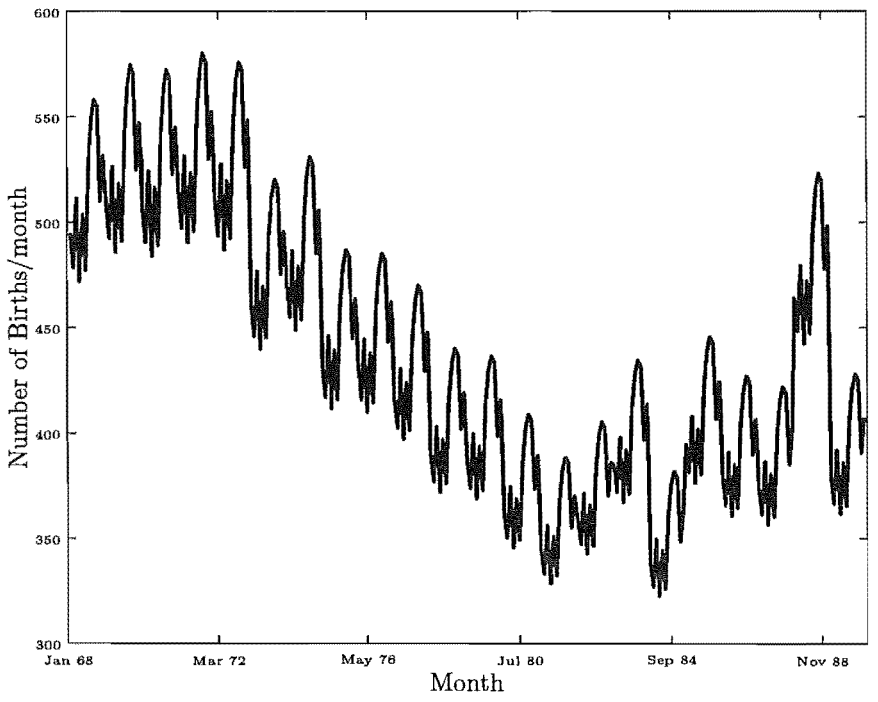


Figure 5.3: Estimated numbers of Canterbury births, 1968—1989.

	m	D	AIC	SC	P
Period 1					
$Temp^{m}_{meanMA30}$	2	166.18	170.18	174.34	48.86
$\sin(\frac{2\pi i}{12}) + \cos(\frac{2\pi i}{12})$	3	165.01	171.01	177.30	45.02
Period 2					
$\sin(\frac{2\pi i}{12}) + \cos(\frac{2\pi i}{12})$	3	717.95	723.95	733.90	203.93
$\sin(\frac{2\pi i}{12}) + \cos(\frac{2\pi i}{12})$ + 2nd harmonics	5	714.74	724.74	741.33	197.95

Table 5.1: Goodness-of-fit statistics for the best two candidate seasonality models, by period (m represents the number of parameters in each model).

NAR will be included into further modelling as a regressor, allowing the dependency of SIDS on the number of infants at risk to be examined. This differs from the notion of an offset (Agresti, 1996), where the baseline population is considered to have a fixed effect on the incidence of the event.

Seasonality

A comprehensive examination of seasonality candidates was performed at a monthly level. Analysis techniques were comparable to those presented in Chapter 4 for the daily case, except that logistic models were replaced by Poisson models (see Chapter 7 for details). The seasonality variables examined were monthly analogues to those detailed in Section 1.8.14 (page 39).

Table 5.1 details goodness-of-fit statistics for the best two candidate seasonality models for each period.

Period 1: The best two candidate seasonality measures for Period 1 were $Temp^{m}_{meanMA30}$ and $\sin(\frac{2\pi i}{12}) + \cos(\frac{2\pi i}{12})$. $Temp^{m}_{meanMA30}$ is the monthly analogue of $Temp_{meanMA30}$, the 30 day restrospective moving average of the mean daily temperature; $\sin(\frac{2\pi i}{12}) + \cos(\frac{2\pi i}{12})$ is a sinusoid function with a period corresponding to 12 months, where $i = 1, \ldots, 59$ in Period 1. Both variables performed similarly with respect to goodness-of-fit statistics, with $Temp^{m}_{meanMA30}$ out performing the sinusoid function with the addition of penalties for added model complexity (AIC and SC). $Temp^{m}_{meanMA30}$ was chosen to model the seasonal profile of SIDS in Period 1. Given that the goodness-of-fit statistics could not separate the two candidates, the simpler functional form (with one less parameter) was taken.

Period 2: As seen in Table 5.1, the best two candiate seasonality measures for Period 2 were the sinusoid, and the sinusoid with second harmonics. As in Period 1, both variables performed similarly with respect to goodness-of-fit statistics, with the simpler first order sinusoid being preferred by the statistics that penalise for added model complexity. The variable defined by $\sin(\frac{2\pi i}{12}) + \cos(\frac{2\pi i}{12})$ was chosen to model the Period 2 seasonal fluctuations in the SIDS profile: this variable was less complex, with two less parameters.

It is interesting to note, that in both Period 1 and Period 2, the choice of monthly seasonality variable was analogous to the variables used to model the seasonal profile at a daily level.

Climate Variables

For the continuous climate variables (temperature, wind speed, north, south east and west wind velocity, humidity, pressure, rainfall, sunshine, radiation and dewpoint) a total of 91 monthly summary variables were created from the continuous daily climate data.

Simple Descriptors

The following basic monthly summaries were created for the continuous climate variables. These incorporate information on daily minimum, mean and maximum measures. In detail the following variables were

created, where X represents the climate variable of interest (for example $X = Temp$, or $X = Pres$):

- $X^m_{mean(mean)}(i) = \text{mean}_{month(i)}(X_{meanday0})$, is the mean over month i of the mean value of X on the day of interest, $day0$;

- $X^m_{mean(min)}(i) = \text{mean}_{month(i)}(X_{minnday0})$, is the mean over month i of the minimum value of X on the day of interest, $day0$;

- $X^m_{mean(max)}(i) = \text{mean}_{month(i)}(X_{maxday0})$, is the mean over month i of the maximum value of X on the day of interest, $day0$;

- $X^m_{min(min)}(i) = \text{min}_{month(i)}(X_{minday0})$, is the minimum over month i of the minimum value of X on the day of interest, $day0$;

- $X^m_{max(max)}(i) = \text{max}_{month(i)}(X_{maxday0})$, is the maximum over month i of the maximum value of X on the day of interest, $day0$.

Variability

A sixth variable is defined which summarises how much variation there has been in X over month i. This is given by:

- $X^m_{mean(std)}(i) = \text{mean}_{month(i)}(X_{stdday0})$, is the mean over month i of the standard deviation of X on the day of interest, $day0$.

Weekly differences

The final two monthly climate variables defined for the continuous climate covariates capture information on the difference between X on the day of interest and over the previous week:

- $X^m_{+diff}(i) = \text{mean}_{month(i)}(X_{diffday0-7} > 0)$, is the mean over month i of all the positive values of $X_{diffday0-7}$;

- $X^m_{-diff}(i) = \text{mean}_{month(i)}(X_{diffday0-7} < 0)$, is the mean over month i of all the negative values of $X_{diffday0-7}$.

These two variables each capture a different profile of $X_{diffday0-7}$: $X^m_{+diff}(i)$ is the monthly average when X is higher on $day0$ than it was over the seven days previously, while $X^m_{-diff}(i)$ is the monthly average when X is lower on $day0$ than it was over the seven days previously.

Component number	General Description	Variance Explained	Cumulative Var. Exp.
PC1	Contrast between East, North, South, Sun against Humid, West	0.406	0.406
PC2	Contrast between Humid, Rain, South, West against Humid(std), West(diff), Rain(diff), South(diff)	0.239	0.645
PC3	Standard deviations against East, Humid, West(−diff), NE	0.149	0.794
PC4	Contrast of difference variables	0.058	0.852
PC5	Weighted mean of Dew, Rad, WindS	0.040	0.892
PC6	Contrast between Pres, E, against West, Temp(+diff), Sun(+diff), NE	0.021	0.913
PC7	Contrast between Rad, WindS against West(−diff), NE	0.017	0.930
PC8	Mean of North(diff), Humid(diff), N, NW, Mixed	0.014	0.944
PC9	Humid	0.011	0.955

Table 5.2: Summary of the monthly climatic components, Period 1.

Wind direction

The categorical daily predominant wind direction variable was summarised at a monthly level by creating a set of ten variables, each representing the percentage of days in month i that each wind direction category predominated (see Table 1.4, page 28 for details of coding). These ten new variables will be denoted as follows:

$$WindD_j^m(i)$$

which corresponds to the % of days in month i where the predominant wind direction was class j. The ten classes, j are: C calm; N north; NE north-east; E east; SE south-east; S south; SW south-west; W west; NW north-west; M no predominant wind direction (mixed).

Monthly Climatic Components

A set of climate components for each period were created using principal component analysis (PCA). Section 4.1 gives full details of the theory behind this analysis. These components were constructed as a way of capturing all the information provided by the 101 monthly variables in a smaller, interpretable dataset.

Period 1: Methods 1, 2 and 3 for choosing the number of components to retain and rotate returned nine, six and nine for each method respectively (see Section 4.1, page 79 for details of the methods used). With two methods agreeing in their results, nine components were retained. Full details of component loadings are given in Appendix E, Table E.1. A summary of the monthly component structure, along with the proportion, and cumulative proportion of variance accounted for by each component, appears in Table 5.2.

The first component, which accounts for approximately 40% of the proportion of variance accounted for, is essentially defined as a contrast of east, north and south wind velocity, sunshine and temperature against humidity and west wind velocity. The second component accounts for a further 24% of the variance. It is again defined by a contrasting structure incorporating rainfall alongside other variables. The third component is predominantly defined by standard deviation summary variables, and accounts for 15% of the variation. The structure of the remaining six components in the set is presented in Table 5.2.

Period 2: Methods 1, 2 and 3 returned eleven, nine and eleven components to be retained and rotated respectively. As with Period 1, two methods were in agreement, therefore eleven components were retained. A complete list of component loadings for the eleven component set is given in Table E.2, Appendix E. Table 5.3 presents a summary of the component structures along with variance details. The first component is a contrast between east and north wind velocity, sunshine and temperature against humidity and west wind velocity, and accounts for 32% of the variance in the original variables. This structure is very similar to that found in the first PC in Period 1. The second component accounts for a further 22% of the variance, and is defined by a weighted mean of dewpoint, radiation, wind speed and west wind velocity variables. With the exception of component 8 and 11, the remaining components have a structure that is defined by varying contrasts between variables. An outline of the structure is given in Table 5.3.

Component number	General Description	Variance Explained	Cumulative Var. Exp.
PC1	Contrast between East, North, South, Temp, Humid(std) against Humid, West	0.322	0.322
PC2	Mean of Dew, Rad, WindS, West	0.222	0.544
PC3	Contrast between Humid, Rain, South, West against Humid(std), North, Sun, Temp	0.118	0.662
PC4	Contrast between Humid, Pres against North, Pres(std), West, Sun	0.068	0.730
PC5	Contrast of Difference variables against WindD	0.050	0.780
PC6	Contrast between South, East, Sun, North, Temp against Humid, Rain	0.035	0.815
PC7	Contrast of Difference variables against WindD	0.031	0.846
PC8	Mean of Rad, WindS	0.021	0.867
PC9	Contrast of minimum variables	0.020	0.885
PC10	Contrast of maximum variables	0.017	0.904
PC11	Pres(min)	0.012	0.915

Table 5.3: Summary of the monthly climatic components, Period 2.

It is interesting to note that the structure of the PC sets created for the daily climate data was defined solely in terms of weighted mean components. In contrast, the PC sets created for both Period 1 and Period 2 monthly climate data were defined by a combination of contrast components and weighted mean components.

The monthly climatic summaries and the PC sets, will be utilised, alongside seasonality and the number of infants at risk of SIDS, in further analysis of the relationship between the incidence of SIDS in Canterbury and climatic patterns. Chapter 6 presents the initial examination of this monthly profile.

Chapter 6

Poisson Regression Analysis of Monthly SIDS Counts

Monthly summaries of SIDS, the population at risk of SIDS, and climate variables were defined in Chapter 5. This chapter utilises these variables and examines the relationship between climate and the incidence of SIDS in Canterbury at the monthly level. The initial distribution used to model the number of SIDS per month is the Poisson distribution, a distribution commonly applied to count data. At a daily level, SIDS outcomes were essentially binary, and were modelled using logistic regression where the dichotomous variable was categorised by days where no SIDS deaths occurred, and days where at least one SIDS occurred. As detailed in Chapter 5 (page 105), the monthly distribution of observed SIDS numbers ranges from zero to nine per month. Thus, the binomial distribution is no longer an appropriate choice to model SIDS at this level, the Poisson distribution is the natural distribution to use. The Poisson distribution can be characterised in terms of the "law of rare events". Cameron & Trivedi (1998) describe this as "the total number of events [that] will follow, approximately, the Poisson distribution if an event may occur in any of a large number of trials but the probability of occurrence in any given trial is small". The number of SIDS per month can therefore be considered to follow a Poisson distribution.

Let y^m denote the number of SIDS per month, with the rate (mean number of SIDS per month) given by λ. The Poisson density is then

$$P(y^m = q) = \frac{\exp(-\lambda)(\lambda)^q}{q!}, \qquad q = 0, 1, 2, \ldots \tag{6.1}$$

with $\mu = \sigma^2 = \lambda$. This characteristic of equal mean and variance in the Poisson distribution is referred to as the equidispersion property (Cameron & Trivedi, 1998).

6.1 Poisson Regression

As stated previously (Chapter 3), the overall aim of this study is to identify the best fitting, yet meteorologically valid and interpretable model to describe the incidence of SIDS in terms of weather patterns, as characterised by one or more climatic variables. The Poisson regression model, a regression method based on the Poisson distribution, has been widely applied in many areas to describe count data, including organisational justice and employee health (Kivimaki et al., 2003); mortality, morbidity and injury in single parent families (Weitoft et al., 2003); and gender differences in melanoma (Nieto et al., 2003). It has also been utilised previously in the analysis of SIDS with respect to climate (Knöbel et al., 1995). Knöbel et al. (1995)

used Poisson regression analysis to relate daily SIDS rates to visibility (utilised as an optimetrical measure of air pollution) and various climate measures including temperature and rainfall. They found that visibility and temperature were significantly associated with the incidence of SIDS.

The general principles of linear regression extend to Poisson regression, as they did with logistic regression (Chapter 3). Many textbooks present detailed coverage of this topic, including Cameron & Trivedi (1998) and Greene (2000). A brief overview of the model and estimation techniques are delineated below.

6.1.1 The Poisson Regression Model

Let y_i^m denote the number of SIDS in month i, $i = 1, \ldots, n$ ($n = 59$ for Period 1, and $n = 204$ for Period 2). Let $\mathbf{y}^m$ represent the vector of monthly responses $[y_1^m, y_2^m, \ldots, y_n^m]'$ and let $\mathbf{X}^m$ denote the matrix of monthly climatic covariates $\mathbf{x}_i^m = [1, x_{1i}^m, x_{2i}^m, \ldots, x_{ki}^m]'$, where there are a total of k covariates.

The Poisson regression model is then defined in terms of the mean (or expected value) of y_i^m, given $\mathbf{x}_i^m$, with the specific form given by

$$\lambda(\mathbf{x}_i^m) = E[y_i^m|\mathbf{x}_i^m] = \exp(\mathbf{x}_i^m\boldsymbol{\beta}), \tag{6.2}$$

where $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_k]'$ corresponds to the vector of regression coefficients. The conditional distribution of the response variable is assumed to be of the form

$$y_i^m|\mathbf{x}_i^m \sim \text{Poisson}[\lambda(\mathbf{x}_i^m)]. \tag{6.3}$$

That is, conditional on $\mathbf{x}_i^m$, y_i^m follows a Poisson distribution with mean $\lambda(\mathbf{x}_i^m)$. The log-linear transformation of $\lambda(\mathbf{x}_i^m)$ is given in equation 6.4 below,

$$\log(\lambda(\mathbf{x}_i^m)) = \mathbf{x}_i^m\boldsymbol{\beta}. \tag{6.4}$$

A major assumption of the Poisson regression model is that

$$E[y_i^m|\mathbf{x}_i^m] = \lambda(\mathbf{x}_i^m) = \exp(\mathbf{x}_i^m\boldsymbol{\beta}) = Var[y_i^m|\mathbf{x}_i^m]. \tag{6.5}$$

This follows from the equidispersion property of the Poisson distribution. The validity of this assumption in the case of modelling the monthly distribution of SIDS will be assessed in the model diagnostic section of the analysis process (Section 6.3.3).

Generalised Linear Models

Both the Poisson regression model and the logistic regression model (Section 3.1.1, page 63) belong to the class of generalised linear models. Generalised linear models (GLMs) form a broad class of models that extend the traditional linear models (McCullagh & Nelder, 1989). All GLMs consist of three components. The first component is a random component, which corresponds to the probability distribution of the response variable. This probability distribution is required to be a member of the exponential family of distributions. The second component is a systematic component, which defines the covariates, or independent variables in the model. These enter the model in the form of a linear combination. The final component of a GLM is the link function. This link function allows the expected value of the response to depend on the linear predictor defined by the systematic component. Mathematically, the GLM is

$$f(E[y]) = \mathbf{X}\boldsymbol{\beta} \tag{6.6}$$

where f is the appropriate link function. In particular, logistic regression consists of a logit link function and a binomial distribution (see equations 3.3 and 3.4, page 63), whereas Poisson regression is specified by a log link function and a Poisson distribution (equations 6.3 and 6.4). Many other widely used statistical models also fit into the class of GLMs, including classical linear regression and ANOVA models.

Fitting the Poisson Regression Model

The likelihood of the Poisson regression model is defined as

$$l(\beta) = \prod_{i=1}^{n} \frac{\exp(-\lambda(\mathbf{x}_i^m))(\lambda(\mathbf{x}_i^m))^{y_i^m}}{y_i^m!}. \tag{6.7}$$

The corresponding log-likelihood equation is given by

$$\begin{aligned} L(\beta) &= \sum_{i=1}^{n} \left[-\lambda(\mathbf{x}_i^m) + y_i^m \log(\lambda(\mathbf{x}_i^m)) - \log(y_i^m!) \right] \\ &= \sum_{i=1}^{n} \left[-\exp(\mathbf{x}_i^m \beta) + y_i^m \mathbf{x}_i^m \beta - \log(y_i^m!) \right]. \end{aligned} \tag{6.8}$$

Maximum likelihood estimation is implemented to calculate the parameters of β via an iterative fitting process (Dobson, 2002). Standard errors, and p-values are calculated for the estimated parameters using the asymptotic normality of the MLEs (Agresti, 1996).

Interpretation of the Coefficients of the Poisson Regression Model

The next step in the analysis process is to interpret the Poisson regression model. This is achieved by examining the effect the parameter estimates have on the response variable. As an exponential relationship exists between $\mathbf{y}^m$ and $\mathbf{X}^m$, an increase in x_j^m $(j = 1, \ldots, k)$ has a multiplicative impact on the expected value of y^m, $\lambda(x_j^m)$. For example, holding other covariates fixed, a unit increase in x_j^m has the following effect on $\lambda(x_j^m)$: $\hat{\lambda}(x_j^m + 1) = \hat{\lambda}(x_j^m)\exp(\hat{\beta}_j)$. More generally, if $\hat{\beta}_j > 0$, then $\exp(\hat{\beta}_j) > 1$ and $\hat{\lambda}(x_j^m)$ will increase with increasing x_j^m. Conversely, if $\hat{\beta}_j < 0$, then $\hat{\lambda}(x_j^m)$ decreases as x_j^m increases.

For large samples, the maximum likelihood estimates of β $(\hat{\beta})$ are approximately normally distributed (Agresti, 1996). A confidence interval for a model parameter, β_j, can then be defined as

$$\hat{\beta}_j \pm z_{\frac{\alpha}{2}} ASE, \tag{6.9}$$

where ASE is the asymptotic standard error of $\hat{\beta}$, and $z_{\frac{\alpha}{2}}$ is the normalised z statistic corresponding to a $1 - \frac{\alpha}{2}$ level.

6.2 Statistics for Model Comparison

Three log-likelihood based statistics, and Pearson's chi-squared statistic are utilised as a basis for deciding between competing Poisson regression models. These are the same model fit statistics used in Chapter 3. The three log-likelihood based statistic are the Deviance statistic (D, defined in equation 3.10 or G, equation 3.11, page 65), Akaike's information criterion (AIC, equation 3.12, page 65), and Schwarz's criterion (SC, equation 3.13, page 65). Pearson's chi-squared statistic is based on the Pearson residual, which for the Poisson regression model is defined as

$$r_i^m = \frac{y_i^m - \hat{\lambda}(\mathbf{x}_i^m)}{\sqrt{\hat{\lambda}(\mathbf{x}_i^m)}}. \tag{6.10}$$

Pearson's chi-squared statistic, P, is then the resulting summary statistic formed by summing the squared Pearson residuals, that is:

$$\begin{aligned} P &= \sum_{i=1}^{n} (r_i^m)^2 \\ &= \sum_{i=1}^{n} \frac{(y_i^m - \hat{\lambda}(\mathbf{x}_i^m))^2}{\hat{\lambda}(\mathbf{x}_i^m)}. \end{aligned} \tag{6.11}$$

6.3 Assessing the Fit of the Poisson Regression Model

After completing the Poisson regression analysis, and having identified the best model to relate SIDS to the monthly climatic covariates, the resulting model needs to be examined for its appropriateness and adequacy in terms of how closely it describes the incidence of SIDS. A similar procedure to that presented in Section 4.3 (page 81), which examined both the overall model fit, and the contribution of the individual observations, is used with the Poisson regression model. In addition, the assumption of equality between the conditional mean and variance, as delineated in equation 6.5, is examined for its validity.

6.3.1 Overall Model Fit

Assessing the overall fit of the model is done via a goodness-of-fit statistic, and an examination of predictive ability, as was the case with the analysis of daily SIDS counts via logistic regression (Chapter 3).

Goodness-of-fit

The hypothesis

H_0: All aspects of the fitted model are correct

against

H_A: Not all aspects of the fitted model are correct

is generally tested using Pearson's chi-squared statistic (equation 6.11). When the fitted values, $\hat{\lambda}(\mathbf{x}_i^m)$, are reasonably large (in general exceeding five), P follows a χ^2-distribution (Cameron & Trivedi, 1998). Unfortunately this was not the case with the distribution of monthly SIDS counts. Therefore, rather than utilising P in a formal hypothesis test setting as defined above, it will be examined in a less formal sense.

If $P > n - k$ (where n is the sample size and k the number of variables in the model), it is usually interpreted as evidence of overdispersion (variance > mean), similarly $P < n - k$ is indicative of underdispersion (variance < mean) (Cameron & Trivedi, 1998). Cameron & Trivedi (1998) note that this interpretation of P assumes that $\lambda(\mathbf{x}_i^m)$ was correctly specified. They suggest that if $P \neq n - k$, the conditional mean may in fact have been misspecified. P is therefore utilised in this sense as an indication of misspecification. Examination of the model for over (or under) dispersion is looked at in Section 6.3.3.

Predictive Ability of the Poisson Regression Model

The predictive ability of the model is examined by comparing the predicted probabilities against the observed frequencies. The response variable, y_i^m, has observed values $0, 1, \ldots, \max(y_i^m)$; denote the observed

frequencies as $\bar{p}_j$, that is $\bar{p}_j$ is the proportion of the observed values with $y_i^m = j$ $(j = 0, 1, \ldots, \max(y_i^m))$. Predicted probabilities are denoted by $\hat{p}_j$ where

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(-\hat{\lambda}(\mathbf{x}_i^m))\hat{\lambda}(\mathbf{x}_i^m)^j}{j!} \qquad j = 0, 1, \ldots, \max(y_i^m). \tag{6.12}$$

Comparing $\bar{p}_j$ with $\hat{p}_j$ highlights values where the model is inclined to over or under estimate the observed distribution.

6.3.2 Residual Diagnostics

The standard definition of a residual for a linear model is simply the difference between observed and fitted values. The definition of a residual in the case of nonlinear models is not unique. Two forms of residuals are utilised in this section of analysis: Pearson residuals and Anscombe residuals.

The Pearson residual was defined previously in equation 6.10, and is the basis of the Pearson chi-squared statistic defined in equation 6.11.

The Anscombe residual is generally defined as the functional transformation of y that returns a distributional form closest to normality. This is then standardised to have zero mean and unit variance (Cameron & Trivedi, 1998). For the case where y follows a Poisson distribution, the function $y^{\frac{2}{3}}$ is closest to normality and the Anscombe residual is given by

$$a_i = \frac{1.5\left[(y_i^m)^{\frac{2}{3}} - (\hat{\lambda}(\mathbf{x}_i^m))^{\frac{2}{3}}\right]}{\left[\hat{\lambda}(\mathbf{x}_i^m)\right]^{\frac{1}{6}}}. \tag{6.13}$$

Residuals Against the Predicted Mean

The second stage of assessing the model fit involves examining the contribution to the model of the individual observations. As previously described in Section 4.3.2, this is achieved by examining several measures against the predicted mean $\hat{\lambda}(\mathbf{x}_i^m)$. An initial plot of the Pearson and Anscombe residuals against $\hat{\lambda}(\mathbf{x}_i^m)$ is examined for any unusual patterns or discrepancies.

Normal Scores Plot

A normal scores plot of the residuals is utilised to examine how close to normality the residuals are. This plot compares the residuals from the fitted model with the predicted residuals under the assumption that they are in fact normally distributed. This plot is generated by plotting the ordered model residuals (r_i^m or a_i) against the expected order statistics of a normal sample. If the residuals do come from a normal distribution, a straight line is produced. Deviations from the linear trend give an indication of lack of model fit over certain values of the dependent variable. The normal scores plot is useful for examining the accuracy of fit of extreme values in the model (McCullagh & Nelder, 1989).

The normal scores plot was not suitable to use in the Logistic regression case (Chapter 3); in Logistic regression it is assumed that the binomial distribution describes the distribution of the errors in the model (Hosmer & Lemeshow, 2000).

Leverage Matrix

A third plot is examined, which assesses the effect of the ith observation. The leverage matrix, H, is defined as

$$H = W^{\frac{1}{2}} X (X'WX)^{-1} X' W^{\frac{1}{2}} \tag{6.14}$$

where $W = \text{Diag}(E[y_i^m|\mathbf{x}_i^m])$ (Cameron & Trivedi, 1998). The ith diagonal entry h_{ii} is then plotted against the observation number. This is the Poisson analogue of equation 4.14 (page 83), which in the logistic

regression analysis was utilised to examine the effect of the deletion of the ith observation on both Pearson's chi-squared statistic and the model parameters. In the Poisson regression scenario, the plot of h_{ii} is examined for any influential observations, identified as being greater than $2n/m$ (m represents the number of parameters in the model).

Autocorrelation Function

The final stage of the residual diagnostics is to examine the autocorrelation function (ACF) (Section 4.3.2, page 84). This is necessary as the monthly series still contains an underlying time ordering. When visually inspecting the ACF of the residuals, any deviations from a purely random series are highlighted by extreme values.

If the ACF shows that no serial correlation exists in the residuals, then the static Poisson regression model (as in equation 6.3) is considered sufficient. A pure Poisson point process generates a series of independent observations (Cameron & Trivedi, 1998) therefore, although the SIDS counts are ordered over time, it is not unusual that the static model is adequate.

In the senario where the ACF indicates serial correlation remaining in the regression residuals, autoregressive covariates $(y^m_{i-1}, y^m_{i-2}, \ldots, y^m_{i-k})$ is added to the model, giving an exponential conditional mean of the following form:

$$E[y^m_i|\mathbf{x}^m_i, y^m_{i-1}, y^m_{i-2}, \ldots, y^m_{i-k}] = \exp(\mathbf{x}^m_i\beta + \rho\mathbf{y}^m_{i-k}), \tag{6.15}$$

where $\mathbf{y}^m_{i-k} = [y^m_{i-1}, y^m_{i-2}, \ldots, y^m_{i-k}]^{prime}$

The maximum likelihood theory is not affected by the addition of $\mathbf{y}^m_{i-k}$ to the set of regressors. Cameron & Trivedi (1998) state that the Poisson estimates remain consistent in the presense of autocorrelation. They show that, if lagged dependent variables are included in the model as regressors, and there is no serial correlation in y^m_i after controlling for the regressors, standard errors also remain consistent. Similar lagged regression coefficients are presented by Zeger & Qaqish (1988) in a Markov model, Aitkin & Alfo (2003) in an extension of transitional GLMs, and Toscas & Faddy (2003) in a longitudinal extended Poisson process model.

Although the model in equation 6.15 is explosive for $\rho > 0$, in all modelling where the additional autoregressive regressor is required $\rho < 0$ (see Results sections for further details).

The ACF of the residuals is used to assess the appropriate AR(k) order required, alongside the mixture transition distribution SC-based test of Raftery & Tavare (1994).

6.3.3 Testing for Overdispersion

A set of data is considered to be overdispersed when the conditional variance exceeds the conditional mean, that is $Var[y^m_i|\mathbf{x}^m_i] > E[y^m_i|\mathbf{x}^m_i]$. A simple indication of the magnitude of overdispersion can be obtained by examining the ratio of the sample variance to the sample mean. If this is greater than unity, overdispersion is a potential problem which may need to be accounted for in the initial model specification. Poisson regression decreases the conditional variance slightly as compared to the sample variance, but overdispersion is likely to remain an issue when $s^2 > 2\bar{x}$ (Cameron & Trivedi, 1998). Tests for overdispersion examine the equidistribution property, with the null hypothesis of equality of the mean and variance considered against the alternative that the variance exceeds the mean, as stated below:

$$\mathrm{H}_0\text{: } E[y^m_i|\mathbf{x}^m_i] = Var[y^m_i|\mathbf{x}^m_i]$$

against

$$\mathrm{H}_A\text{: } E[y^m_i|\mathbf{x}^m_i] < Var[y^m_i|\mathbf{x}^m_i].$$

Two test statistics for examining the overdispersion hypothesis are utilised, as presented by Dean (1992). The first test statistic, denoted OD_A, assumes $E[y^m_i|\mathbf{x}^m_i] \approx \lambda(\mathbf{x}^m_i)$ with $Var[y^m_i|\mathbf{x}^m_i] \approx \lambda(\mathbf{x}^m_i)(1 + \tau\lambda(\mathbf{x}^m_i))$

(where τ is small). This is a simple extension to the log-linear Poisson model presented in Section 6.1.1. The test statistic is

$$OD_A = \frac{\sum_{i=1}^{n} \left[(y_i^m - \hat{\lambda}(\mathbf{x}_i^m))^2 - (1 - \hat{h}_{ii})\hat{\lambda}(\mathbf{x}_i^m)\right]}{\sqrt{2\sum_{i=1}^{n} \hat{\lambda}(\mathbf{x}_i^m)^2)}} \tag{6.16}$$

where h_{ii} is the ith diagonal element of H, defined in equation 6.14.

The second statistic used to test for overdispersion is OD_B. This uses a simple variance inflation: $Var[y_i^m|\mathbf{x}_i^m] = \lambda(\mathbf{x}_i^m)(1+\tau)$, giving

$$OD_B = \frac{1}{\sqrt{2n}} \sum_{i=1}^{n} \left[\frac{(y_i^m - \hat{\lambda}(\mathbf{x}_i^m))^2 - y_i^m + \hat{h}_{ii}\hat{\lambda}(\mathbf{x}_i^m)}{\hat{\lambda}(\mathbf{x}_i^m)}\right]. \tag{6.17}$$

Dean & Lawless (1989) show that the distribution of the two test statistics, OD_A and OD_B, is asymptotically normal.

6.4 Statistical Modelling Methods

The initial analysis step involves examining the effect of the variable describing the population at risk, over and above seasonality. As described in Chapter 5, the monthly variable used to describe seasonality in Period 1 is $Temp^m_{meanMA30}$, the mean of $Temp_{meanMA30}$ over each calendar month. In Period 2, seasonality is described using a sinusoid function, composed of sine and cosine curves (with a period of twelve to represent the annual cycles in the seasonal component of the distribution of SIDS).

Models are constructed using the following model forms.

Baseline model

(1) SIDS $\propto$ intercept + season

(2) SIDS $\propto$ intercept + season + number of infants at risk (NAR)

Climate model

(3) SIDS $\propto$ intercept + season + NAR + climate variables.

A baseline model is formed, selecting the best model form from (1) or (2) above. These baseline forms are also examined without the intercept term. The additional contribution of the monthly climate variables (Model (3) above) is initially examined individually, over and above the baseline model. After identifying potentially significant climatic covariates, multivariate Poisson regression models are formed by combining multiple climate variables into Model (3). These will be compared with the regression models formed from the PC sets defined in Chapter 5. Finally, the model identified as best describing the relationship between climate and monthly SIDS numbers will be assessed for its appropriateness and adequacy. This modelling process is completed separately for Period 1 (1968—1972) and Period 2 (1973—1989). Note: where necessary, climate variables are deseasoned using the procedure outlined in Section 3.3.2, page 68.

The Poisson regression analyses were performed using SAS (*SAS (R) Proprietary Software Release (8.01.01)*) through the GENMOD procedure. An $\alpha \leqslant 0.05$ level of significance is used throughout this analysis procedure.

6.5 Results

Figure 6.1 shows the relationship between NAR and the intercept when modelling SIDS per month (for Period 2). It is interesting to note that in all models that include both the intercept and NAR, the regression parameter corresponding to NAR is negative (see modelling results in this chapter and also Chapters 6 and

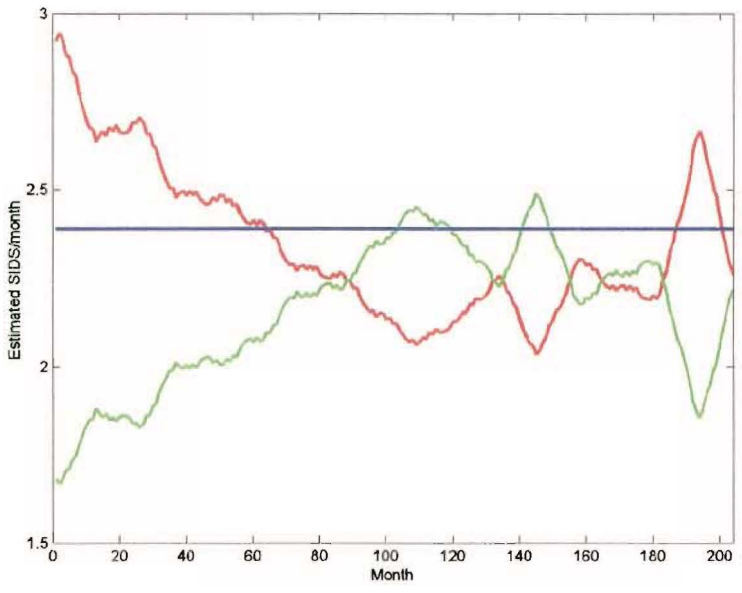


Figure 6.1: Relationship between NAR and intercept when using Poisson regression to model monthly SIDS counts (Period 2); $SIDSintercept + NAR$; $SIDSNAR$; $SIDSintercept$.

8). This may be because NAR changed very little, relative to its absolute size, over the period, and so is nearly aliased with the constant term.

This contradicts with what is the expected relationship between NAR and the incidence of SIDS; as the number of infants at risk of SIDS increases, so should the number of SIDS. Figure 6.1 shows the intercept only model corresponding to the mean number of SIDS per month during Period 2 (2.4), with the NAR and $int + NAR$ models essentially inverses, with a reflection around approximately 2.2. The estimated number of SIDS with the NAR only model, follows the same general pattern of NAR over time. Models were therefore considered involving only the intercept or NAR, as well as both. Most models presented for SIDS at a monthly level contained both intercept and NAR, but for completeness and comparison, where models within the main body of the thesis are presented without both intercept and NAR, the corresponding model containing both intercept and NAR are presented in Appendix G.

6.5.1 Period 1 (1968—1973)

Table 6.1 presents four candidate baseline Poisson regression models for Period 1, with parameter estimates and corresponding goodness-of-fit statistics. The models are denoted P1(1), P1(2), P1(3) and P1(4). Both P1(1) and P1(3) are based solely on $Temp^{m}_{meanMA30}$, whereas P1(2) and P1(4) include the additional variable, NAR. P1(3) and P1(4) have the same model form as P1(1) and P1(2) respectively, but without the constant (or intercept) term in the model.

P1(1) has the lowest values across the three log-likelihood based statistics (D, AIC and SC) of the four models. P1(2), in turn, returns the lowest value of Pearson's chi-squared statistic, but both the constant term and NAR lack significance. Table 6.1 also presents the rank sum value (calculated using the method described in Chapter 3). These values show that P1(1), with the lowest rank sum of 5, performs best out

Model	Variables	$\hat{\beta}$ (*se*)	D	AIC	SC	P	Rank sum
P1(1)	Intercept	0.955 (0.342)	166.182	170.182	174.337	48.862	5
	$Temp^{m}_{meanMA30}$	-0.059 (0.030)					
P1(2)	Intercept	10.331 (8.172)	166.664	172.664	178.896	46.646	9
	$Temp^{m}_{meanMA30}$	-0.058 (0.030)					
	NAR	-0.0016 (0.001)					
P1(3)	$Temp^{m}_{meanMA30}$	0.017 (0.009)	173.330	175.330	177.407	65.429	14
P1(4)	$Temp^{m}_{meanMA30}$	-0.058 (0.030)	170.264	174.264	178.419	49.338	12
	NAR	0.0002 (0.0001)					

Table 6.1: Baseline model results for Period 1, Poisson regression.

of the four candidate baseline models. P1(1), consisting of an intercept and a seasonality term, is therefore taken as a baseline model in the next analysis step.

P1(4) is also used as a baseline model in the next stage of the analysis. P1(4) contains NAR alongside seasonality. Although P1(4) did not perform as well as P1(1) with respect to the goodness-of-fit statistics, the information contained in the NAR term (which describes the number of infants at risk of SIDS in any month) was considered potentially useful in terms of describing the risk of SIDS, and therefore this model was considered further.

The next stage of the Poisson regression analyses involved incorporating each of the climatic variables into the baseline model (with both P1(1) and P1(4) examined). No significant relationships were identified over and above the baseline models of seasonality, or seasonality and NAR.

Similarly, non-significant effects resulted when examining the contribution of the nine principal components defined in Chapter 5, via PC regression. These components effectively summarise the climate information for Period 1, but with no individual climatic variables significantly related to the incidence of SIDS, over and above either baseline model (P1(1) or P1(4)), it was expected that no relationship was found between the PCs and the incidence of SIDS.

The two baseline models, P1(1) and P1(4), will therefore be considered further in terms of model adequacy and predictive ability.

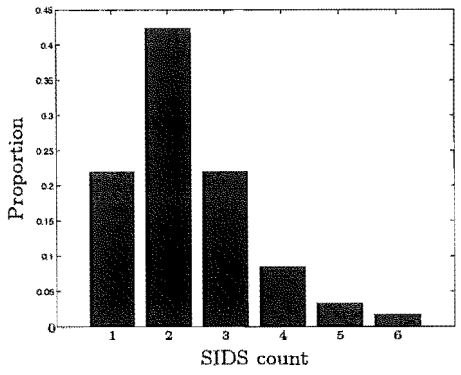
Assessing the Fit of the Baseline Poisson Regression Models

Assessing the fit of the Period 1 baseline models P1(1) and P1(4) (detailed in Table 6.1) falls into three basic categories: overall model fit, residual diagnostics, and testing for overdispersion in the data. Notationally, variables relating to P1(1) will contain a subscript '1', and correspondingly, variables relating to P1(4) will contain a subscript '4'.

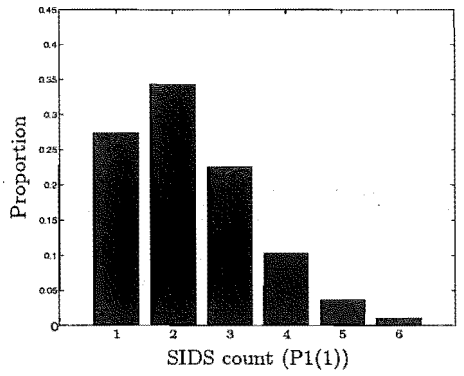
Overall Model Fit of the Baseline Poisson Regression Models

Pearson's chi-squared statistics for P1(1) and P1(4) were $P_1 = 48.862$ and $P_4 = 49.338$ respectively. As the fitted values $\hat{\lambda}_i(\mathbf{X}^m)$ $(i = 1, 4)$ are reasonably small (ranging from approximately 0.9 to 1.9 SIDS per month) P_i does not follow the standard χ^2 distribution. Both P1(1) and P1(4) have $df = 57$, $(df = n - m = 59 - 2 = 57)$, approximately 16% larger than P_1 or P_4. This may be indicative of underdispersion in the model.

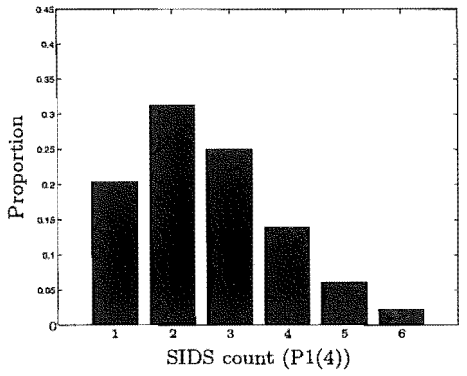
The observed probability distribution of the number of SIDS per month is presented in Figure 6.2, alongside the predicted probability distributions for both P1(1) and P1(4). The distributions of P1(1) (Figure 6.2 (b)) and P1(4) (Figure 6.2 (c)) have slight differences: the predicted proportion of months with no SIDS deaths, and months where one SIDS occurred, is slightly larger in the P1(1) model. In contrast,



(a) Observed proportion of SIDS ($\bar{p}$)



(b) Predicted proportion of SIDS ($\hat{p}$, P1(1))



(c) Predicted proportion of SIDS ($\hat{p}$, P1(4))

Figure 6.2: Observed and predicted proportions of SIDS counts, Period 1 (predicted proportions calculated baseline models P1(1) and P1(4)).

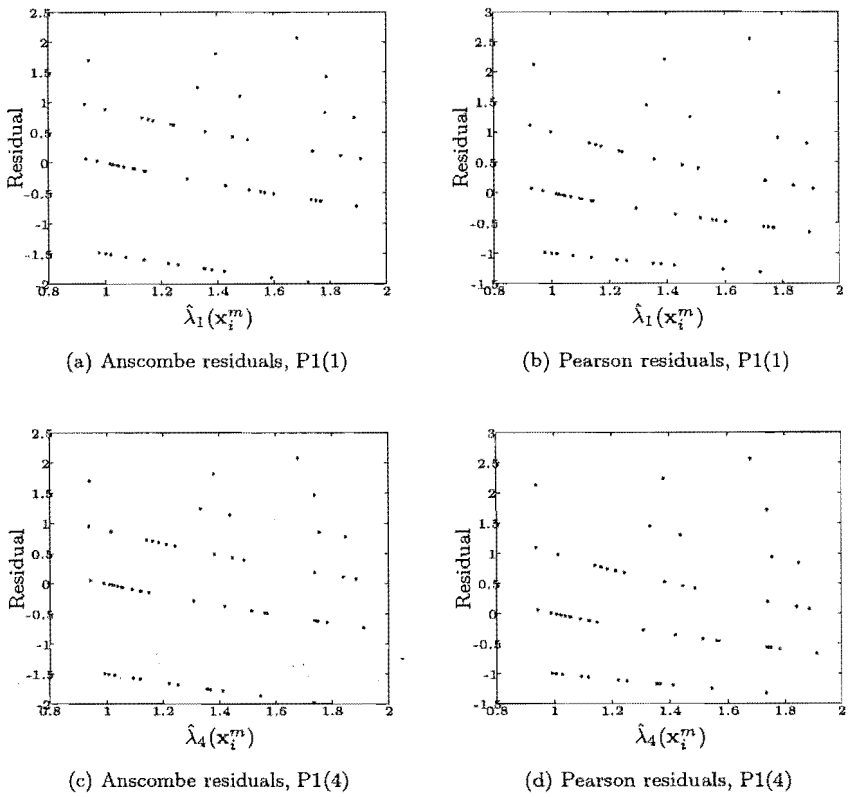
the predicted proportion of months with two or three SIDS is higher in the P1(4) model. Compared to the observed distribution of SIDS, both P1(1) and P1(4) fail to predict the large number of months with one SIDS death.

Residual Diagnostics for the Baseline Poisson Regression Models

Figure 6.3 shows both the Anscombe and Pearson residuals from P1(1) and P1(4) plotted against the predicted mean $\hat{\lambda}(\mathbf{X}^m)$. Figure 6.3 (a) and (b) correspond to the P1(1) residuals, while Figure 6.3 (c) and (d) show the P1(4) residuals. All four plots show the same trend: a banding effect, corresponding to an increase in the predicted mean. Visually, it is difficult to identify values that have been poorly fitted.

Normal score plots for the Anscombe and Pearson residuals from P1(1) and P1(4) are presented in Figure 6.4. The P1(1) residuals appear in Figure 6.4 (a) and (b), with the residuals from P1(4) shown in Figure 6.4 (c) and (d). The normal score plots for both residuals types, and both models, appear similar. The linear trend is poorly followed, especially at the lower range of values. The higher residual values also seem to deviate from the line of normality, but with less residuals in this range, it is not as easily seen. This suggests that the models (P1(1) and P1(4)) may not be estimating $\hat{\lambda}(\mathbf{X}^m)$ successfully.

Figure 6.5 presents the diagonal entries of the leverage matrix ($\hat{h}_{ii}$) defined in equation 6.14, for both



(a) Anscombe residuals, P1(1)

(b) Pearson residuals, P1(1)

(c) Anscombe residuals, P1(4)

(d) Pearson residuals, P1(4)

Figure 6.3: Predicted mean against resduals, Period 1 (P1(1) and P1(4)).

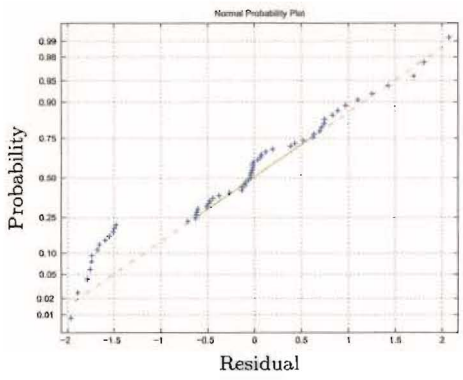
P1(1) and P1(4). Four potential outliers are highlighted by ◯. The bound $2m/n = 0.068$ is shown by – – – – – on both plots. The three largest values of $\hat{h}_{ii}$ for each of P1(1) (Figure 6.5 (a)) and P1(4) (Figure 6.5 (b)) fall above this bound.

The three months identified as potential outliers are June 1968, June 1969 and July 1972. The corresponding values of $Temp^m_{meanMA30}$ and NAR for these three observations are given in Table 6.2. The three potential outliers are all winter months, with one to three SIDS occurring. The $Temp^m_{meanMA30}$ values are similar across the three months. There are no trends evident that may have resulted in these months being outliers. Removing any points from the dataset is not an option as this would corrupt the underlying time series inherent in the data.

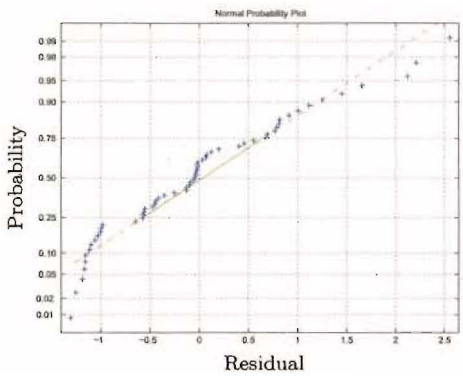
Figure 6.6 presents the autocorrelation functions (ACFs), for lags of up to 15 months, for the Pearson residuals of P1(1) and P1(4). The ACFs show no evidence of serial correlation in either the P1(1) or P1(4)

	June 1968	June 1969	July 1972
# SIDS	3	2	1
Season	Winter	Winter	Winter
$Temp^m_{meanMA30}$	5.44	5.24	5.38
NAR	5708	5764	5896

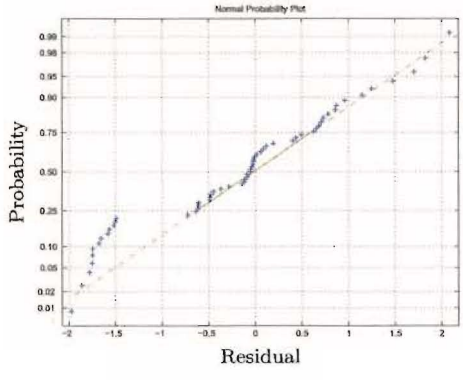
Table 6.2: Covariate values for outlier months in baseline models, Period 1.



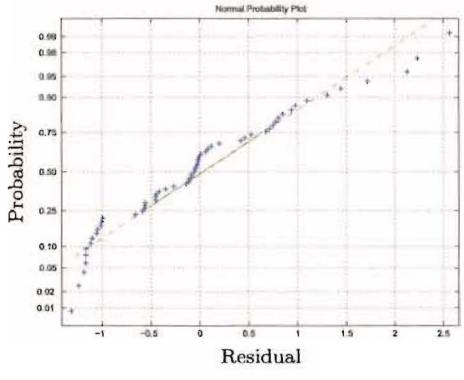
(a) Anscombe residuals, P1(1)



(b) Pearson residuals, P1(1)

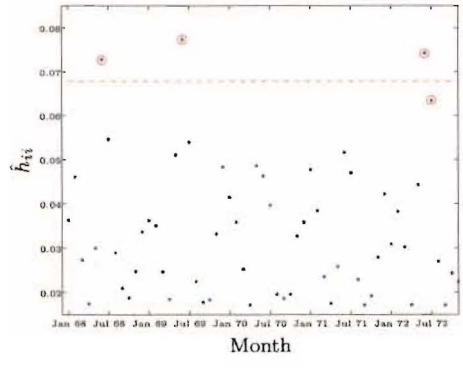


(c) Anscombe residuals, P1(4)

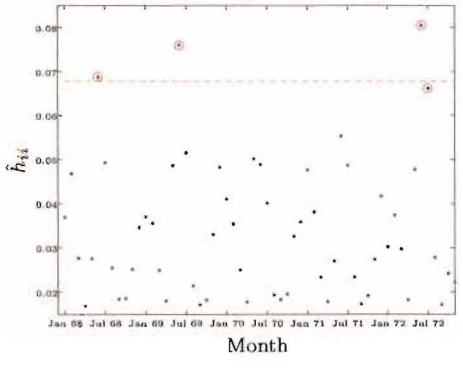


(d) Pearson residuals, P1(4)

Figure 6.4: Normal scores plot to assess the normality of residuals, Period 1 (P1(1) and P1(4)).

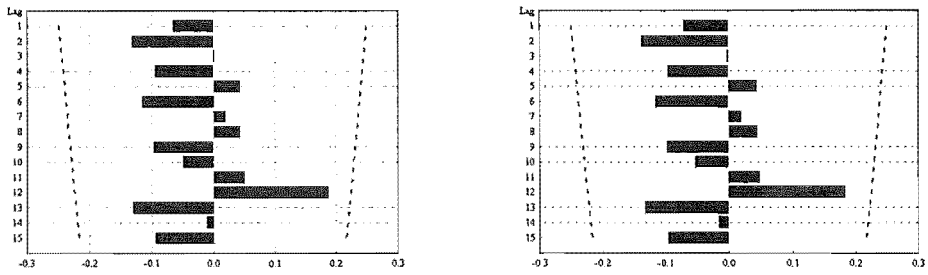


(a) P1(1)



(b) P1(4)

Figure 6.5: Leverage matrix over time, Period 1, (P1(1) and P1(4)).



(a) Pearson residuals, P1(1)

(b) Pearson residuals, P1(4)

Figure 6.6: Autocorrelation function of the residuals, Period 1 (P1(1) and P1(4)).

model, with the series appearing purely random. The ACFs of the Anscombe residuals also do not show any evidence of serial correlation (not pictured).

Testing for Overdispersion in the Baseline Poisson Regression Models

The final stage of assessing the fit of the Period 1 baseline models P1(1) and P1(4) involved testing for overdispersion. The ratio of the observed variance (1.26) to the mean number of SIDS per month (1.34) is 0.944. This value is slightly smaller than one, indicating that underdispersion may be evident in the distribution of SIDS in Period 1. This underdispersion may no longer be an issue after the Poisson modelling process, but it is not expected that significant overdispersion will be identified. The overdispersion statistics, defined in equations 6.16 and 6.17 are presented below for completeness:

$$\begin{aligned} OD_{A1} &= 0.584 & OD_{B1} &= -0.763 \\ OD_{A4} &= -0.544 & OD_{B4} &= -0.728 \end{aligned} \tag{6.18}$$

With p-values ranging from 0.30 to 0.34, these four statistics give no indication of overdispersion in the baseline models.

The overall performance of the Period 1 baseline models, P1(1) and P1(4) was not good. In both cases, the models failed to predict the large proportion of months where one SIDS occurred (Figure 6.2), and normal score plots (Figure 6.4) also gave an indication of lack of model fit. There was evidence of underdispersion seen in the models, highlighted when comparing Pearson's chi-squared statistics to the corresponding model degrees of freedom, and also in the sample variance to mean ratio.

Both baseline models, P1(1) ($intercept + Temp^{m}_{meanMA30}$) and P1(4) ($Temp^{m}_{meanMA30} + NAR$) appeared to be similar in terms of residual diagnostics and dispersion tests.

Interpretation of the Baseline Poisson Regression Models

The Period 1 baseline model P1(4) = $Temp^{m}_{meanMA30} + NAR$ will be utilised to briefly examine the influence of the covariates of the predicted number of SIDS. Although the overall performance of this model is lacking, it is of interest to see the effect of the covariates. The parameter estimates corresponding to the covariates show a decreased SIDS risk, with increasing $Temp^{m}_{meanMA30}$, highlighting the seasonal trend underlying the SIDS distribution. A one degree increase in the monthly average of $Temp_{meanMA30}$ corresponds to a 6% decrease in the number of SIDS (95%CI: 0%—11%). In contrast, an increase in the number of infants at

risk of SIDS (NAR) corresponds to an increase in the risk of SIDS, specifically, every additional 100 infants at risk of SIDS in a month, increases the SIDS risk by 2% (95%CI: 0%—4%).

6.5.2 Period 2 (1973—1989)

The four candidate baseline Poisson regression models for Period 2 are presented in Table ??, with parameter estimates and corresponding goodness-of-fit variables. The models are denoted P2(1), P2(2), P2(3) and P2(4) where P2(1) + NAR = P2(2), and similarly P2(3) + NAR = P2(4). P2(3) and P2(4) have the same model form as P2(1) and P2(2) respectively, but do not contain an intercept (or constant) term.

P2(1) (intercept + seasonality) has the lowest values, of the four models, across the three log-likelihood based goodness-of-fit statistics. This model ranks second with respect to Pearson's chi-squared statistic, and with a final rank sum of five, rates best across the four goodness-of-fit statistics utilised. Adding NAR to P2(1) gives P2(2), which rated second on the log-likelihood statistics, and first with respect to P. Overall, P2(2) had the second lowest rank sum value. The contribution of the additional information made to the model by NAR in terms of describing the underlying trends in the population at risk was considered an important consideration when modelling the incidence of SIDS. Therefore P2(2) was taken as the baseline model for further analysis of the Period 2 data.

It is interesting to note the poor performance of P2(3), the model containing only sinusoids, with no constant term. The addition of NAR to P2(3), giving P2(4), brings the model to a much more comparable level with P2(1) or P2(2). In this sense, NAR is essentially acting as a surrogate for the constant term in P2(4). P2(4) will therefore also be incorporated as a baseline model in the analysis process. The overall effect of both a constant term and NAR, as in P2(2), may swamp the model, making it difficult to identify climatic effects. Utilising P2(4) as a baseline model may aid in the process of teasing out potential climatic relationships with SIDS.

Univariate Climate Models

Baseline Model P2(2):

$$\log(\lambda(\mathbf{x}^m)) = \beta_0 + \beta_1 \sin(\frac{2\pi t}{12}) + \beta_2 \cos(\frac{2\pi t}{12}) + \beta_3 NAR \tag{6.19}$$

The model presented in equation 6.19 above (P2(2)) was taken as the baseline model for this step in the analysis procedure, where the effect of adding additional climate variables to the baseline form is examined. Initially, each climate variable was added individually to the model to enable potentially significant climatic relationships to be identified, over and above the baseline model.

Only two climate variables, out of the possible 101, were identified as being significantly related to the incidence of SIDS, over and above the relationship between seasonality and the population at risk described in the baseline model (P2(2)). These were $WindS^m_{mean(std)}$ and $South^m_{-diff}$. Table ?? presents details for these two models, including parameter estimates and goodness-of-fit statistics. They are denoted Model 1 and Model 2 corresponding to the incorporation of $WindS^m_{mean(std)}$ and $South^m_{-diff}$ respectively. Model 1 is marginally better than Model 2, with respect to the goodness-of-fit statistics.

Figure ?? illustrates the estimated effect of $WindS^m_{mean(std)}$ on SIDS risk, by season, while the remaining covariate values were held fixed at their seasonal mean value. The estimated mean number of SIDS, $\hat{\lambda}(\mathbf{X}^m)$, is plotted against $WindS^m_{mean(std)}$ for each season. A unit increase in $WindS^m_{mean(std)}$ has a multiplicative impact on the risk of SIDS. Therefore, for every one knot increase in $WindS^m_{mean(std)}$, the estimated mean number of SIDS per month increases by 18%. Figure ?? shows that winter has the highest risk of SIDS, ranging from 3.1 to 4.9 SIDS per month as estimated by Model 1. The predicted SIDS risk for spring and autumn is similar, with spring SIDS risk ranging from 1.9 to 3.0 SIDS per month and for autumn from 1.7

$Humid^m_{mean(max)}$	$Pres^m_{mean(mean)}$	$Pres^m_{mean(min)}$	$Pres^m_{mean(max)}$
$West^m_{mean(mean)}$	$Pres^m_{min(min)}$	$Dew^m_{max(max)}$	$Pres^m_{max(max)}$
$West^m_{max(max)}$	$WindS^m_{-diff}$	$WindS^m_{+diff}$	
$North^m_{+diff}$	Dew^m_{+diff}	$WindD^{md}_{NW}$	

Table 6.3: Climate variables significantly assiciated with SIDS, over and above the baseline model formed by P2(4), Period 2 ('d' corresponds to deseasoned variables).

to 2.7 SIDS per month. As expected with the underlying seasonal pattern in SIDS, summer has the lowest risk, ranging from 1.0 to 1.6 SIDS per month over the range of $WindS^m_{mean(std)}$ values plotted. SIDS risk appears linear over the range of $WindS^m_{mean(std)}$ values. Similar trends follow from Model 2 (not shown).

No significant relationships between any of the principal components defined in Chapter 5 and the incidence of SIDS in Period 2, where P2(2) was taken as the baseline model form, were identified.

Baseline Model P2(4):

$$\log(\lambda(\mathbf{x}^m)) = \beta_1 \sin(\frac{2\pi t}{12}) + \beta_2 \cos(\frac{2\pi t}{12}) + \beta_3 NAR \tag{6.20}$$

A total of fourteen monthly climate variables, out of the possible 101, were found to be significantly related to the incidence of SIDS, over and above the baseline model given in equation 6.20 above (P2(4)). Full model details, including parameter estimates and goodness-of-fit statistics are presented in Table F.1; a list of the significant covariates is given in Table 6.3. Of the fourteen variables, $WindD^m_{NW}$ is the only variable that was significantly associated with season, and therefore deseasoned. This is denoted by a 'd'.

Of the univariate models, with baseline form P2(4), the model which incorporated $Humid^m_{mean(max)}$ attained the lowest goodness-of-fit statistics. This model implies the mean number of SIDS per month increased by 2% for every 1% increase in $Humid^m_{mean(max)}$. Figure 6.7 illustrates the estimated effect of $Humid^m_{mean(max)}$ on the predicted SIDS numbers, by season. This again, highlights winter as season of highest SIDS risk, and combined with high values of $Humid^m_{mean(max)}$, the estimated number of SIDS per month reaches 3.6. As shown in Table F.1, $Pres^m_{max(max)}$ also rated well on the goodness-of-fit statistics.

No significant relationships were identified between any of the principal components defined in Chapter 5 and the incidence of SIDS in Period 2, where P2(4) was taken as the baseline model form.

Overall, the univariate models containing an intercept term, alongside seasonality and risk population variables (baseline P2(1)), performed better with respect to the goodness-of-fit statistics compared to models without an intercept term (baseline P2(4)). This is expected with the better performance of the underlying baseline model P2(2) as compared to P2(4). More climate variables were found to be significantly associated with SIDS risk, over and above the baseline model for P2(4) as compared to P2(2).

Multiple Climatic Models

Baseline Model P2(2):

The two variables related to SIDS at a univariate level, $WindS^m_{mean(std)}$ and $South^m_{-diff}$, were incorporated together into a multiple climatic model. The resulting model, denoted Model 3, is presented in Table 6.4. The addition of $South^m_{-diff}$ to Model 1, producing Model 3, significantly reduced the log-likelihood of the Model 1 ($G = 5.291, p = 0.021$).

Baseline Model P2(4):

Table 6.4 details three further multiple climatic models, corresponding to a baseline model form given by P2(4) (Seasonality + NAR). These three models, denoted Model 4, 5 and 6, each contain two climate

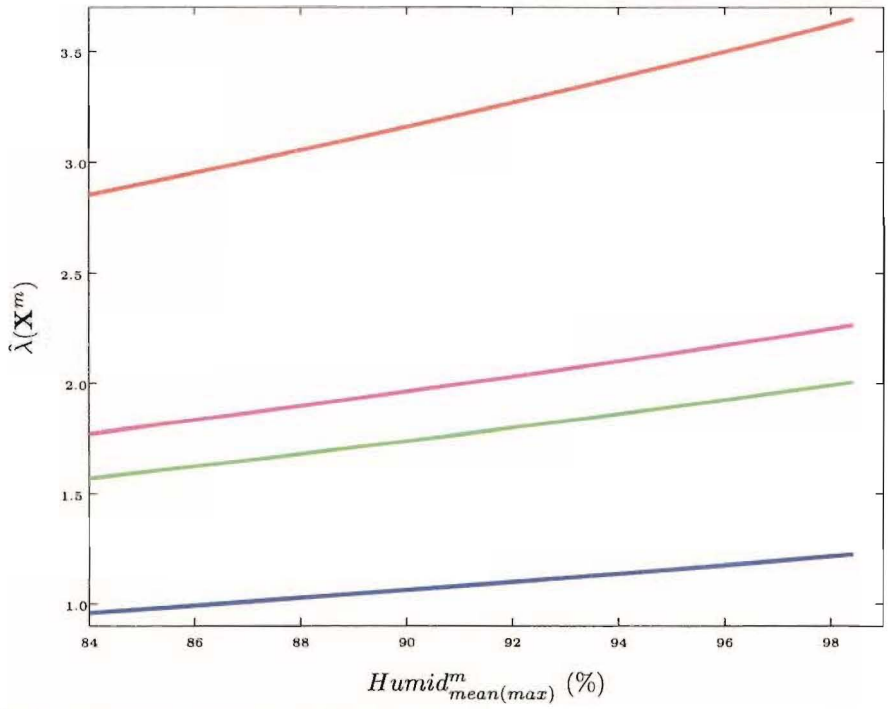


Figure 6.7: Univariate climate model highlighting the effect of $Humid^m_{mean(max)}$ on the predicted mean rate of SIDS, by season, baseline model P2(4), Period 2 (— Summer, — Autumn, — Winter, — Spring).

variables, $WindD^{md}_{NW}$ alongside $West^m_{mean(mean)}$, $Dew^m_{max(max)}$ or $West^m_{max(max)}$. There is little difference between these three models, and they are all outperformed by Model 3 (containing the constant term). Model 3 is therefore assessed in terms of model adequacy and overall fit. The underlying climatic dependencies described by this model are then examined in a detailed model interpretation.

Assessing the Fit of Model 3

As described in Section 6.3, assessing the fit of Model 3 falls into three basic categories, overall model fit, residual diagnostics, and testing for overdispersion in the data. Dependent on the results of this process, Model 3 can then be used to aid in understanding the climatic dependency of SIDS.

Overall Model Fit for Model 3

Pearson's chi-squared statistic for Model 3 was $P = 189.249$. As discussed in Section 6.3, as the fitted values $\hat{\lambda}(\mathbf{X}^m_i)$ are reasonably small (ranging from 1.0 to 5.0 SIDS per month), P does not follow the standard χ^2 distribution. Degrees of freedom for Model 3 are $df = 198$, slightly larger than P. Informally, this gives no obvious indication that the conditional mean may have been misspecified.

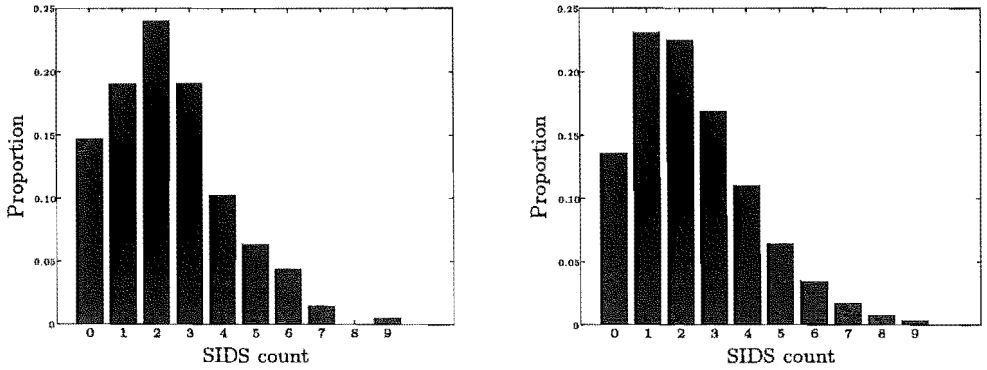
Figure 6.8 presents both the observed and predicted probability distribution of SIDS counts. Both plots present similar right-skewed distributions, although the predicted distribution ($\hat{p}$, Figure 6.8 (b)) appears bi-modal, with a twin peak at one and two. In contrast, the observed distribution of SIDS counts ($\bar{p}$, Figure 6.8 (a)) has a single mode at two SIDS per month. The class of one SIDS per month appears to be the only category where Model 3 overestimates the observed distribution.

Residual Diagnostics for Model 3

The residuals for Model 3 are plotted against the predicted mean, $\hat{\lambda}(\mathbf{X}^m)$ in Figure 6.9. Figure 6.9 (a) shows the Pearson residuals, while Figure 6.9 (b) presents the Anscombe residuals. Both plots show a similar

Model	Variable	$\hat{\beta}$ (se)	D	AIC	SC	P
Model 3	Intercept	1.274 (0.527)	705.604	717.604	737.512	189.249
	$sin(\frac{2\pi t}{12})$	-0.342 (0.068)				
	$cos(\frac{2\pi t}{12})$	-0.441 (0.067)				
	NAR	-0.0002 (0.0001)				
	$WindS^{m}_{mean(std)}$	0.164 (0.072)				
	$South^{m}_{-diff}$	0.165 (0.078)				
Model 4	$sin(\frac{2\pi t}{12})$	-0.337 (0.067)	720.131	730.131	746.722	217.378
	$cos(\frac{2\pi t}{12})$	-0.445 (0.067)				
	NAR	0.0001 (0.0001)				
	$West^{m}_{mean(mean)}$	0.114 (0.061)				
	$WindD^{md}_{NW}$	1.578 (0.737)				
Model 5	$sin(\frac{2\pi t}{12})$	-0.336 (0.067)	719.461	729.461	746.052	216.649
	$cos(\frac{2\pi t}{12})$	-0.446 (0.067)				
	NAR	0.0001 (0.0001)				
	$Dew^{m}_{max(max)}$	0.036 (0.018)				
	$WindD^{md}_{NW}$	1.683 (0.744)				
Model 6	$sin(\frac{2\pi t}{12})$	-0.335 (0.067)	721.291	731.291	747.812	214.313
	$cos(\frac{2\pi t}{12})$	-0.444 (0.067)				
	NAR	0.0001 (0.0001)				
	$West^{m}_{max(max)}$	0.048 (0.024)				
	$WindD^{md}_{NW}$	1.636 (0.746)				

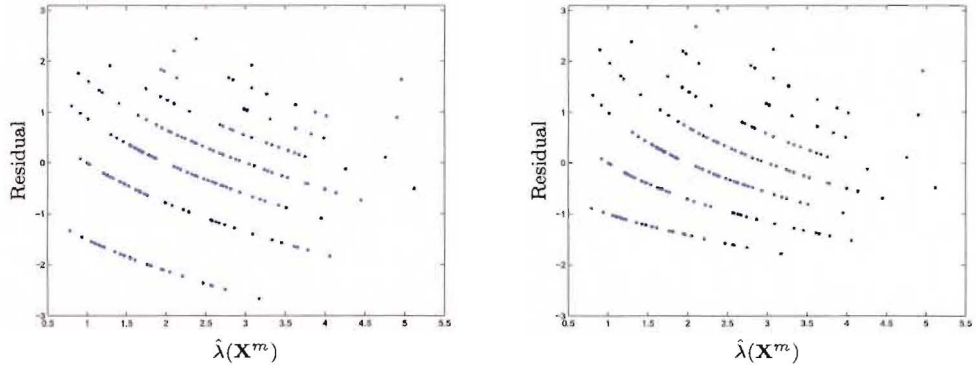
Table 6.4: Details for various multivariate climate models, Period 2 ('d' corresponds to deseasoned variables).



(a) Observed proportion of SIDS ($\bar{p}$)

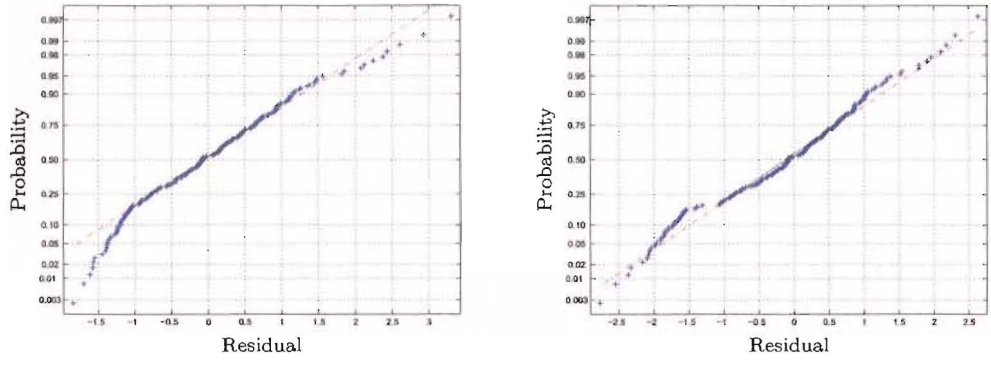
(b) Predicted proportion of SIDS ($\hat{p}$)

Figure 6.8: Observed and predicted proportions of SIDS counts, Period 2, Model 3.



(a) Anscombe residuals (b) Pearson residuals

Figure 6.9: Predicted mean against residuals, Period 2, Model 3.



(a) Pearson residuals (b) Anscombe residuals

Figure 6.10: Normal scores plot to assess the normality of residuals from Model 3, Period 2.

residual pattern, with the Pearson residuals recording slightly lower values. Visually, it is difficult to identify poorly fit values.

Normal score plots are presented in Figure 6.10 (a) (Pearson residuals) and Figure 6.10 (b) (Anscombe residuals). The relationship appears linear for the mid-range residual values, but deviates from the line of normality at the extremes of the residual range. This pattern is seen in both the Pearson and Anscombe residuals, and suggests that Model 3 may not be successfully estimating $\hat{\lambda}(\mathbf{X}^m)$ for months where there were either no SIDS, or where large numbers of SIDS occurred.

The leverage matrix, defined in equation 6.14, is useful in highlighting influential observations. Figure 6.11 presents the diagonal entries of this leverage matrix ($\hat{h}_{ii}$) plotted in over time. This shows eight outlying observations, highlighted by ○. The bound $2m/n$ corresponds to a value of 0.059 for Model 3, and is shown on Figure 6.11 by – – – – –. Covariate details of the outlying observations are given in Table 6.5, and show that the eight months were either winter or spring months. Recorded SIDS numbers for ranged from zero to eight per month for the outlying points. No trends were evident that may have resulted in these months being outliers. As a result of the time series nature of the data, removal of any of these eight points from the

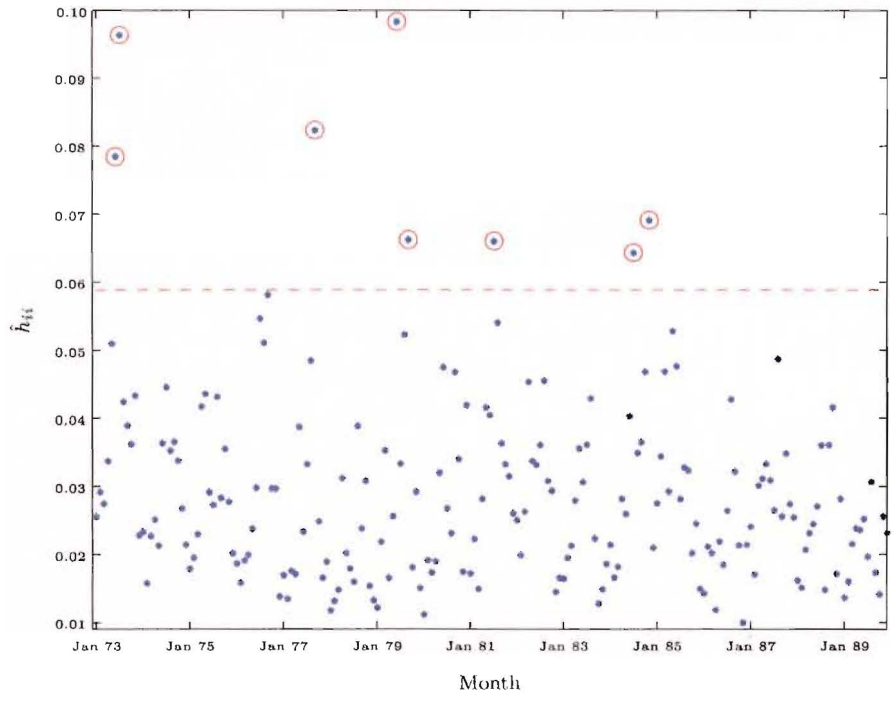


Figure 6.11: Leverage matrix for each observation, Period 2, Model 3.

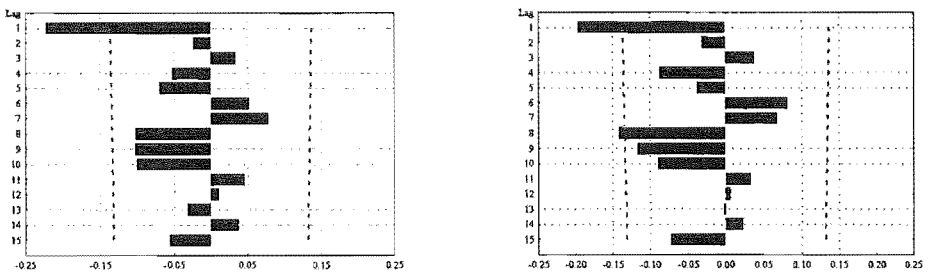
dataset is not an option, as this would interfere with the regular time-ordered nature of the observations.

Figure 6.12 presents the autocorrelation functions for the Pearson (Figure 6.12 (a)) and Anscombe (Figure 6.12 (b)) residuals from Model 3, for lags of up to 15 months. Both these autocorrelation functions (or correlograms) show a large value at lag$(t-1)$, implying a possible short-term correlation remaining in the residuals of Model 3. An autoregressive covariate, comprised of lagged SIDS counts, was therefore incorporated into Model 3. This was a way of accounting for serial correlation in the model, allowing for more accurate interpretation of the climatic profile of SIDS risk.

Details of the model containing the autoregressive covariate given by the number of SIDS per month, lagged by one month (denoted y^m_{i-1}), is presented in Table 6.6. This model, which is defined as Model 3 + y^m_{i-1}, is denoted Model 7. Inclusion of y^m_{i-1} significantly reduced the log-likelihood ($G = 3.668$, $p = 0.055$), which is reflected in lower values of three goodness-of-fit statistics as compared with Model 3. The SC statistic,

Date	Jun 73	Jul 73	Sep 77	Jun 79	Sep 79	Jul 81	Jul 84	Nov 84
Season	Winter	Winter	Spring	Winter	Spring	Winter	Winter	Spring
# SIDS	3	4	0	7	5	2	4	4
$sin(\frac{2\pi t}{12})$	0	-0.5	-1	0	-1	-0.5	-0.5	-0.5
$cos(\frac{2\pi t}{12})$	-1	-0.8666	0	-1	0	-0.866	-0.866	0.866
NAR	5649.2	5626.5	4788.6	4452.3	4406.4	4079.2	4195.5	3936.2
$WindS^m_{mean(std)}$	5.206	5.777	4.504	5.475	5.500	3.029	4.156	5.777
$South^m_{-diff}$	-1.043	1.516	-4.696	-0.907	-2.804	-1.540	-3.148	-1.199

Table 6.5: Covariate values for potentially influential observations, Model 3, Period 2.



(a) Pearson residuals

(b) Anscombe residuals

Figure 6.12: Autocorrelation function of the residuals of Model 3, Period 2.

Variable	$\hat{\beta}$ (se)	95% CI	D	AIC	SC	P
Intercept	1.706 (0.54)	0.647 - 2.765	701.936	715.936	739.163	178.404
$sin(\frac{2\pi t}{12})$	-0.453 (0.074)	-0.599 - -0.307				
$cos(\frac{2\pi t}{12})$	-0.502 (0.070)	-0.638 - -0.365				
NAR	-0.0002 (0.0001)	-0.0004 - -0.0000				
$WindS^{m}_{mean(std)}$	0.158 (0.073)	0.016 - 0.300				
$South^{m}_{-diff}$	0.180 (0.080)	0.027 - 0.334				
y^{m}_{i-1}	-0.099 (0.030)	-0.157 - -0.041				

Table 6.6: Detailed model descriptors of Model 3 + autoregressive term (Model 7), Period 2.

which penalises harshly for the extra parameter now in the model, is slightly larger than the SC statistic for Model 3. Figure 6.13 presents the autocorrelation functions for the residuals from Model 7. There appears to be a large ACF value at lag$(t-9)$, but this is not considered extreme. Overall, the ACFs do not show any evidence of a non-random pattern in the time series of residuals from Model 7.

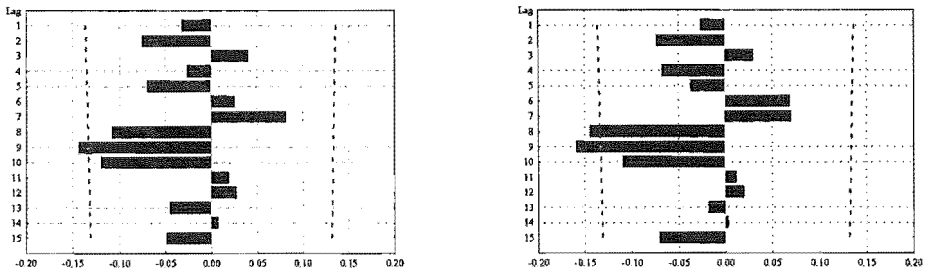
Testing for Overdispersion in Model 3 and Model 7

The final stage of assessing the fit of the model involves testing the data for overdispersion. Model 3 and Model 7 are examined for any evidence of overdispersion. The ratio of the observed variance to the mean is 1.313. This value is larger than one, indicating that overdispersion may be a problem. As $s^2 = 3.135$ and $\bar{x} = 2.387$, this may no longer be an issue after the Poisson regression modelling process. Values for the two test statistics described in Section 6.3.3, equations 6.16 and 6.17, were as follows:

$$\begin{aligned} OD_{A3} &= -0.500 \qquad & OD_{B3} &= -0.396 \\ OD_{A7} &= -1.058 \qquad & OD_{B7} &= -0.892 \end{aligned} \tag{6.21}$$

With p-values ranging from 0.23 to 0.37, these four statistics given no indication of overdispersion in the Period 2 models. This implies that there is insufficient evidence to show that $E[y^{m}_{i}|\mathbf{x}^{m}_{i}] < Var[y^{m}_{i}|\mathbf{x}^{m}_{i}]$ for either model, implying that the Poisson distribution assumption is valid.

Overall, Model 7 containing $WindS^{m}_{mean(std)}$, $South^{m}_{-diff}$ and y^{m}_{i-1}, appears to adequately describe the relationship between climate and the incidence of SIDS at a monthly level in Period 2. Residual analyses showed some evidence of overestimated predicted SIDS for months where one SIDS occurred, and possible lack of fit at extreme values of SIDS incidence. There was no suggestion of overdispersion in the model.



(a) Pearson residuals

(b) Anscombe residuals

Figure 6.13: Autocorrelation function of the residuals of Model 7, Period 2.

Model Interpretation

The final model chosen to describe the relationship between climate and the monthly number of SIDS in Period 2 was Model 7, which is detailed in Table 6.6 (parameter estimates, standard errors and corresponding confidence intervals are given). Of the six covariates incorporated into the model, two are climatically based, namely $WindS^{m}_{mean(std)}$ and $South^{m}_{-diff}$. The remaining variables describe the underlying seasonality in the model ($\sin(\frac{2\pi t}{12})$ and $\cos(\frac{2\pi t}{12})$), the effect of the population at risk of SIDS (NAR), and the underlying autoregressive structure resulting from the time series nature of the data (y^{m}_{i-1}).

Figure 6.14 presents the log of the estimated mean number of SIDS per month, where $\log[\hat{\lambda}(\mathbf{X}^m)]$ was calculated using the mean values of each covariate for each season. Error bounds were determined using the upper and lower 95% confidence limits of the parameter estimates. This figure shows that Model 7 clearly captures the underlying seasonal distribution of the incidence of SIDS. The predicted number of SIDS for winter months was 4.34 (95% CI: 0.46—40.79), more than three times that of summer where the predicted number of SIDS was 1.31 per month (0.09—18.46). Spring and autumn have similar estimates.

There was a positive effect of both the climatic covariates in the model on the predicted number of SIDS, with a unit increase in $WindS^{m}_{mean(std)}$ corresponding to a 17% increase in the number of SIDS per month (2%—35%) (holding other covariates fixed). Similarly a unit increase in $South^{m}_{-diff}$ results in a 20% increase in the estimated mean number of SIDS (3%—40%). For example, an increase in the mean over a month of the daily standard deviation of wind speed ($WindS^{m}_{mean(std)}$) from four to five knots would result in the estimated mean number of SIDS increasing from 2.63 to 3.08 SIDS per month (when holding the remaining covariates fixed at their overall mean value). The variable $South^{m}_{-diff}$ describes the strength of southern wind velocity on $day0$ as compared to what it was over the previous seven days, averaged for the month over days where $South_{meanday0}$ was less than the retrospective weekly average. An increase in $South^{m}_{-diff}$ implies that the difference between $South_{meanday0}$ and the average over the seven days prior to $day0$ decreased. Increasing $South^{m}_{-diff}$ from -2 to -1 knots results in the estimated number of SIDS increasing from 2.80 to 3.36 SIDS per month, when the remaining covariates are held fixed at their overall mean values. These climatic relationships are illustrated in Figure 6.15, which shows $\hat{\lambda}(\mathbf{x}^m)$ increasing with both increasing $South^{m}_{-diff}$ and $WindS^{m}_{mean(std)}$. Over the range of the independent climatic variables plotted, this relationship appears linear.

Figure 6.16 presents the predicted number of SIDS per month, overlaying the observed numbers for Model 7. The model appears to follow the underlying seasonal trend evident in the observed series. The plot shows the predicted mean number of SIDS per month, so, by nature of the variable, is restricted from

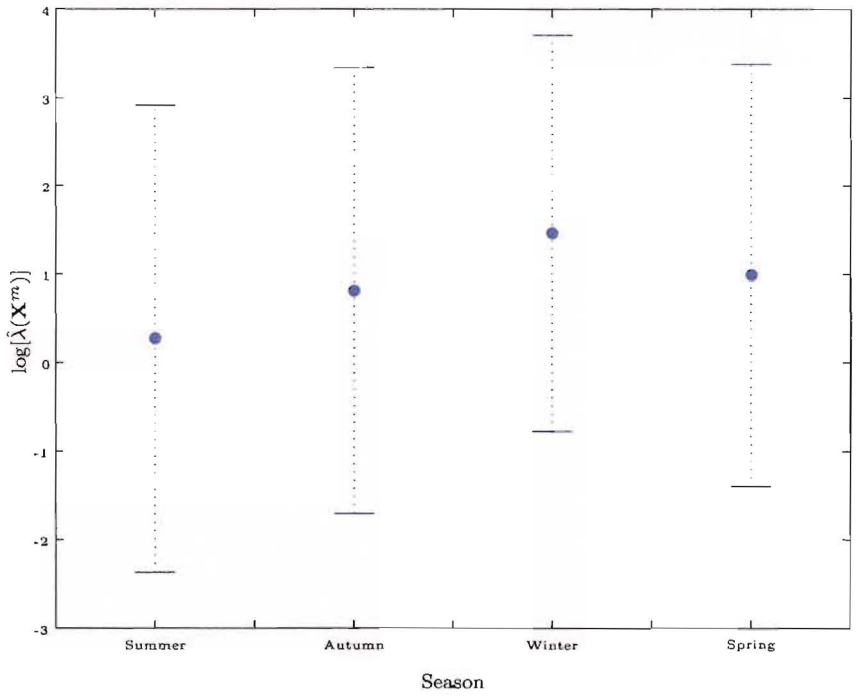


Figure 6.14: Seasonal variation in the log of the estimated number of SIDS (with 95% confidence interval bounds), Model 7, Period 2.

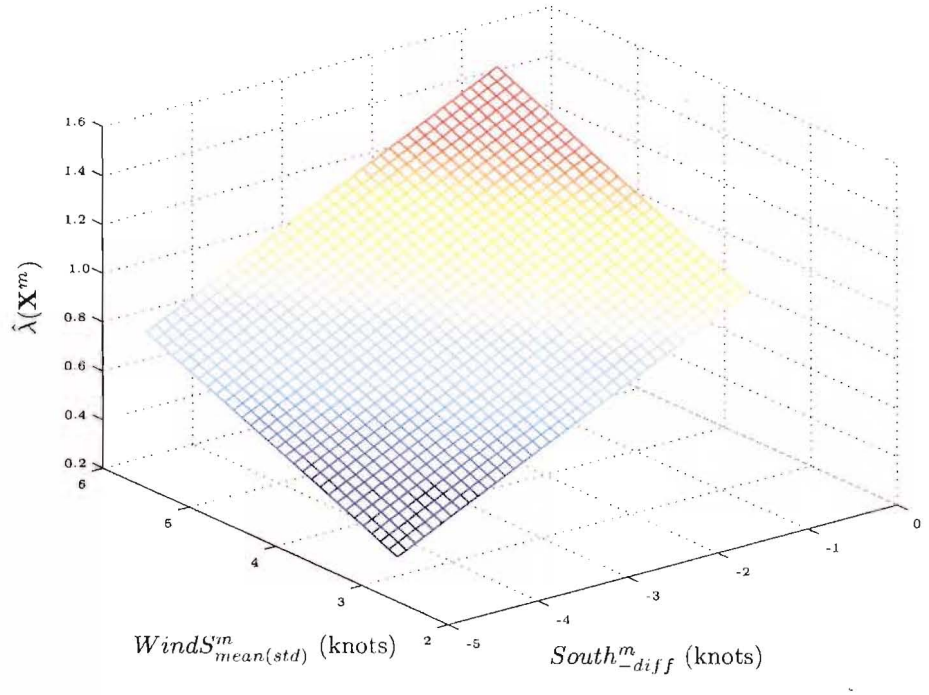


Figure 6.15: Predicted rate of SIDS for varying values of the climatic covariates, Model 7, Period 2.

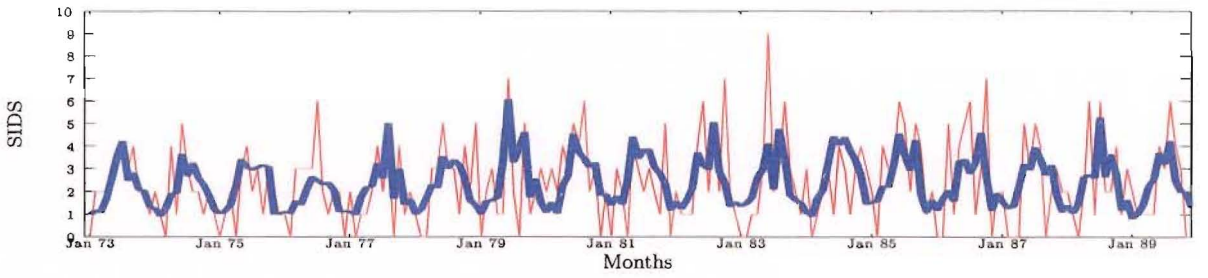


Figure 6.16: Observed —— and pedicted ▬ number of SIDS per month for Model 7, Period 2.

following the extremes of the observed series. The expected number of SIDS does not fall below one death per month, and the Poisson regression model also fails to predict the peaks in the SIDS series. When the predicted series appears to follow the peaks in the observed number of SIDS well, the months with low numbers of SIDS are poorly predicted, an example of this occurs at June 1979.

6.6 Summary and Conclusion

This chapter systematically examined the relationship between climate and the incidence of SIDS at a monthly level via Poisson regression techniques. Poisson regression is suited to modelling data where the dependent variable consists of non-negative integers, or counts, as is the situation with the number of SIDS per month. The model building process involved incorporating a variable describing the number of infants at risk of SIDS in any month, alongside the seasonality function, to form a baseline model. The impact of monthly climate covariates was then examined, over and above the baseline model. This process was completed separately for Period 1 (1968—1972) and Period 2 (1973—1989).

Results for Period 1 failed to find any climatic covariates that were significantly related to the incidence of SIDS, over and above the baseline model. The baseline model did not perform well, with regard to predictive ability and residual diagnostics. Underdispersion in the Period 1 SIDS distribution may have been a factor.

Two baseline model forms were utilised in Period 2; both forms contained seasonality and NAR (the number of infants in the risk population for SIDS) with only one baseline model including an intercept term. The omission of the intercept allowed further climatic patterns to be identified.

Over the two baseline model forms, sixteen climate variables, out of the 101 examined, were found to be significantly related to the incidence of SIDS, over and above the corresponding baseline model. Using the baseline form which included an intercept term, two climate variables were identified: the first, $WindS^{m}_{mean(std)}$, captured the profile of the variability in daily wind speed over a month, while the second variable was a summary of the days over a month where the southern component of wind velocity was less than it had been over the previous seven days ($South^{m}_{-diff}$). Both these variables had a positive effect on the estimated mean rate of SIDS, implying that an increase in either variable resulted in an increased SIDS risk. Similarly, most of the remaining fourteen climate variables, identified as significantly associated with SIDS, corresponded to a positive effect on the estimated mean number of SIDS per month.

Overall, the model containing both $WindS^{m}_{mean(std)}$ and $South^{m}_{-diff}$ adequately described the climatic relationship with SIDS in Period 2. The model appeared to overestimate the observed distribution for the class of one SIDS per month and there was some suggestion that the model may not be successfully estimating the mean monthly SIDS rate for extreme values of the SIDS distribution (no SIDS or large numbers of SIDS per month). An autoregressive component was added to the model to compensate for serial correlation which remained in the residuals, due to the underlying time series nature of the data. Significant overdispersion in

the model was not identified.

It is of interest to note that the climate based monthly principal components defined in Chapter 5 were not significantly related to the incidence of SIDS, over and above the baseline model, in either Period 1 or Period 2. This contrasts with the results in Chapter 4, where detailed daily climate PCs perform well when modelling the incidence of SIDS.

The model presented in this chapter as the best multivariate Poisson regression model to describe the climatic dependency of the incidence of SIDS in Period 2 was not perfect. The next chapter looks at various other model forms that may better capture the profile of the climatic dependency. These include zero-inflated Poisson and hurdle models, which allow for extra zeros within the distribution of SIDS counts, as well as models based mixtures of Poisson distributions and extended Poisson processes.

Chapter 7

Poisson Mixture Modelling of Monthly SIDS Counts

This chapter looks at five different Poisson based mixture model formulations. The Poisson mixture models are able to highlight differential climatic effects between months where SIDS occur and months where there are no SIDS deaths, while retaining the underlying Poisson formulation. Utilising models of this type may better capture the climatic dependency of the monthly incidence of SIDS, enabling an enhanced understanding of the relationship between climate and SIDS deaths. Mixture modelling in this form is uncommon in the general time series framework.

The analysis presented in this chapter is a major extension of that given by Dalrymple et al. (2003), which was the first study to apply mixture based models to a temporal sequence of SIDS counts. Dalrymple et al. (2003) examined only three mixture models in relation to Canterbury monthly SIDS counts, for the restricted time frame of Period 2 (1973—1989), and with a limited climatic focus (humidity and temperature based climate variables). The analysis presented in this chapter extends the work of Dalrymple et al. (2003) in three ways: firstly, two further Poisson based mixture models are presented (negative binomial and extended Poisson process models); secondly, the analysis examines the underlying climatic relationships evident in SIDS counts in both Period 1 and Period 2; and thirdly, a complete climatic profile, involving the 101 climatic covariates defined in Chapter 5 is examined for significant influences on the incidence of SIDS.

The five model formulations presented in this chapter are negative binomial (Greene, 2000), finite mixture (Wang et al., 1998), zero-inflated Poisson (Lambert, 1992), hurdle (Cameron & Trivedi, 1998), and extended Poisson processes (Faddy, 1997). They are all extensions of the Poisson distribution formulation, yet characterise the underlying distribution of SIDS counts in slightly different ways. For example, the finite mixture model views the distribution of SIDS as arising from distinct Poisson distributions, whereas the zero-inflated Poisson model views the distribution as arising from two states, a perfect state where no SIDS occur and an imperfect state where SIDS are described by a Poisson distribution. All of the five formulations can induce overdispersion into the models, while the extended Poisson process formulation is the only method which has the ability of incorporating underdispersion with respect to the Poisson distribution. This dispersion may be a result of natural heterogeneity in the underlying groupings of the observed data, or may in fact arise from the essence of the process producing the slight excess of zeros seen in Period 2. This chapter extends these five model formulations to incorporate the time series nature of the SIDS series. This is achieved by including a lagged response variable as a regressor when residual autocorrelation is present.

Chapter 6 examined the relationship between SIDS and climate at a monthly level using standard Poisson regression techniques. Though the resulting models were considered adequate overall, some suggestion of

lack of fit was highlighted for both months where no SIDS occurred and months with a high SIDS incidence. This was highlighted in both the residual analyses and the need for the inclusion of an autoregressive term in the Period 2 models.

In Chapter 6, the best Poisson regression model predicted only 11% of months with no SIDS deaths in Period 2, whereas, in fact there were 15% of months where no deaths occurred. Although this is not a large distributional difference, and tests in Chapter 6 showed that the model was not overdispersed, the Poisson mixture models may highlight differential climatic dependencies for months of varying SIDS levels. In contrast, the best Poisson regression (baseline form, P1(1), Table 6.1, page 120) for Period 1 overestimated the proportion of months where no SIDS deaths occurred at 27%, when there were only 22% of months with no deaths.

Details for each of the model formulations are given in the following section, alongside the underlying interpretations of the models, how the model parameters are estimated, and an overall explanation of the resulting fitted model.

7.1 Poisson Mixture Model Theory

Poisson mixture models allow a more refined analysis, compared to Poisson regression, and result in a more comprehensive understanding of the relationship between SIDS and climate. Insights into some fundamental and latent structure may also result from Poisson mixture modelling.

A description of the underlying distributions in the five mixture formulations is presented in Table 7.1. In addition, Table 7.1 presents an overview of particular situations where the given models reduce to the Poisson regression model, or one of the other mixture models. This table shows where the Poisson distribution fits into the underlying model formulations. For example, the negative binomial (NB) model is defined by a mixture of a Poisson and gamma distribution. Both the finite mixture (FM) and zero-inflated Poisson (ZIP) models involve Poisson distributions to describe the rate of SIDS. In contrast, in the hurdle model this component is defined by a truncated (at zero) Poisson distribution. A binomial distribution characterises the probability of a zero or non-zero outcome in both the ZIP and the hurdle models. In the FM model, component membership is defined by a multinomial distribution, which reduces to the binomial form when there are two Poisson components in the mixture model.

Four of the five model formulations directly incorporate the Poisson distribution, for certain parameter values, namely, the NB, FM, ZIP and extended Poisson process (EPP) models. The EPP model also encompasses the NB, for a particular set of parameter values. The ZIP model formulation can be considered a sub-class of the family of FM models.

The following sections detail the underlying model formulations for the five methods. Parameter estimation methods are also presented.

7.1.1 Negative Binomial Model

The requirement of equality of the conditional mean and variance in the Poisson regression model can be rather restrictive. A common alternative is the negative binomial (NB) model which allows overdispersion to enter the model through flexible modelling of the variance.

There are many ways the negative binomial distribution can arise (see for example Boswell & Patil (1970)). The most common form is that the observed distribution is Poisson, but the true mean is not perfectly observed. This is incorporated into the model through an unobserved individual heterogeneity which follows a gamma distribution. Numerous texts briefly discuss the negative binomial model, including Dobson (2002), and Burr (1974); both Cooper & Weekes (1983), and Cameron & Trivedi (1998) discuss the

Model		Mixing formulation	Distribution	Particular cases for certain parameter values
Negative binomial	(NB)	Poisson + gamma	$y_i \sim$ negative binomial	• reduces to Poisson
Finite mixture	(FM)	Poisson + multinomial	$y_i \sim$ Poisson(λ_j) with probability p_j	• reduces to ZIP • reduces to Poisson
Zero-inflated Poisson	(ZIP)	Poisson + binomial	$y_i = 0$, with prob $1-p$ $y_i \sim$ Poisson(λ), with prob p	• reduces to Poisson
Hurdle		truncated Poisson + binomial	$y_i = 0$, with prob $1-p$ $y_i \sim$ truncated Poisson, with prob p	
Extended Poisson process	(EPP)	Poisson process	$Y(t)$ = prob distn of SIDS counts	• reduces to Poisson • reduces to NB

Table 7.1: Overview of mixture models.

model in more detail. The following formulation is presented in terms of a mixture distribution, and is an overview of that presented by both Cameron & Trivedi (1998) and Greene (2000).

Let y_i^m denote the number of SIDS in month i, $i = 1, \ldots, n$ ($n = 59$ for Period 1 and $n = 204$ for Period 2); $\mathbf{y}^m$ is then the vector of monthly responses. Let $\mathbf{X}^m$ denote the matrix of monthly covariates with $\mathbf{x}_i^m = [1, x_{1i}^m, \ldots, x_{gi}^m]'$, where there are a total of g covariates.

The Poisson regression model is generalised to the NB model with the addition of an unobserved effect, ϵ_i, to the conditional mean $\mu(\mathbf{x}_i^m)$,

$$\begin{aligned} \log(\mu(\mathbf{x}_i^m)) &= \mathbf{x}_i^m \boldsymbol{\beta} + \epsilon_i \\ &= \log(\lambda(\mathbf{x}_i^m)) + \log(\nu_i), \end{aligned} \tag{7.1}$$

where $\lambda(\mathbf{x}_i^m) = \exp(\mathbf{x}_i^m \boldsymbol{\beta})$. The number of SIDS counts per month, y_i^m, conditional on $\mathbf{x}_i^m$ and ν_i (that is ϵ_i), follows a Poisson distribution, given by

$$P(y_i^m = q | \mathbf{x}_i^m, \nu_i) = \frac{\exp(-\lambda(\mathbf{x}_i^m)\nu_i)(\lambda(\mathbf{x}_i^m)\nu_i)^q}{q!} \qquad q = 0, 1, 2, \ldots. \tag{7.2}$$

The unconditional distribution (with respect to ν_i) of y_i^m is then found by integrating out ν_i; that is

$$P(y_i^m = q | \mathbf{x}_i^m) = \int P(y_i^m = q | \mathbf{x}_i^m, \nu_i) g(\nu_i) d\nu_i \qquad q = 0, 1, 2, \ldots, \tag{7.3}$$

where $g(\nu_i)$ represents a mixing distribution. Equation 7.3 is the expected value of $P(y_i^m | \mathbf{x}_i^m, \nu_i)$ (over ν_i). The unconditional distribution, $P(y_i^m | \mathbf{x}_i^m)$ is defined by the choice of the mixing distribution $g(\nu_i)$. In this case, $g(\nu_i)$ is assumed to follow a one parameter gamma distribution of the form

$$g(\nu_i) = \frac{\theta^\theta}{\Gamma(\theta)} \exp(-\theta\nu_i) \nu_i^{\theta-1}. \tag{7.4}$$

The probability density function for y_i^m is then given by

$$P(y_i^m = q|\mathbf{x}_i^m) = \int \frac{\exp(-\lambda(\mathbf{x}_i^m)\nu_i)(\lambda(\mathbf{x}_i^m)\nu_i)^q}{q!} \frac{\theta^\theta \nu_i^{\theta-1}\exp(-\theta\nu_i)}{\Gamma(\theta)} d\nu_i$$
$$= \frac{\Gamma(\theta+q)}{\Gamma(q+1)\Gamma(\theta)} r_i^q (1-r_i)^\theta \qquad q = 0, 1, 2, \ldots, \tag{7.5}$$

where $r_i = \frac{\lambda(\mathbf{x}_i^m)}{\lambda(\mathbf{x}_i^m)+\theta}$. This form of the NB distribution has a conditional mean and variance given by:

$$E[y_i^m|\mathbf{x}_i^m] = \lambda(\mathbf{x}_i^m)$$
$$Var[y_i^m|\mathbf{x}_i^m] = \lambda(\mathbf{x}_i^m) + \frac{\lambda(\mathbf{x}_i^m)^2}{\theta}, \qquad (\theta > 0). \tag{7.6}$$

It can be seen from equation 7.6 that $Var[y_i^m|\mathbf{x}_i^m] > E[y_i^m|\mathbf{x}_i^m]$. This results in an overdispersed model formulation, with respect to the Poisson distribution.

The canonical link function for the form of the negative binomial distribution modelled in this analysis, is the log-linear transformation (as given in equation 7.1).

Fitting the Negative Binomial Model

Noting that $\frac{\Gamma(y+a)}{\Gamma(a)} = \prod_{j=0}^{y-1}(j+a)$, if y is an integer, gives

$$\log\left(\frac{\Gamma(\theta+y_i^m)}{\Gamma(\theta)}\right) = \sum_{j=0}^{y_i^m-1} \log(j-\theta). \tag{7.7}$$

Substituting this result into equation 7.5 gives the log-likelihood function for the negative binomial distribution, characterised as a mixture of Poisson and gamma distributions:

$$L_{NB}(\theta, \boldsymbol{\beta}) = \sum_{i=1}^{n} \Bigg(\Big[\sum_{j=0}^{y_i^m-1} \log(j-\theta) \Big] - \log y_i^m!$$
$$- (y_i^m + \theta)\log(1 + \frac{1}{\theta}\exp(\mathbf{x}_i^m\boldsymbol{\beta})) + y_i \log\theta + y_i\mathbf{x}_i^m\boldsymbol{\beta} \Bigg) \tag{7.8}$$

The dispersion parameter $\frac{1}{\theta} = d$ is estimated, along with the parameter estimates $\boldsymbol{\beta}$, by maximum likelihood methods, utilising an iterative fitting procedure (Dobson, 2002). Standard errors are computed utilising the asymptotic normality of the maximum likelihood estimators. NB modelling is completed using SAS (*SAS (R) Proprietary Software Release (8.01.01)*), implementing the GENMOD procedure.

The negative binomial distribution reduces to the Poisson density when $d = 0$. See Cameron & Trivedi (1998, page 75) for a proof of this result.

7.1.2 Finite Mixture Model

Finite mixture (FM) models have become increasingly popular in the analysis of a wide range of data. FM models were introduced in Chapter 2 as a way of identifying change points in the SIDS series. This section of work extends those FM models to a class of Poisson mixture models which incorporate covariates into both the mixing proportions, and underlying Poisson distribution. The covariate adjusted mixture model was presented by Schlattmann et al. (1996), in a disease mapping context. Wang et al. (1996) presented two medical examples of mixed Poisson regression models with covariate dependent rates, namely seizure frequency and Ames salmonella assay data.

The FM model assumes an underlying partition of the population into k homogeneous components. Each component has a different SIDS risk level ($\lambda_j(\mathbf{x}_i^m)$), dependent on possibly different covariates. This is a general model which allows mixing with respect to both zeros and positives.

Denote the ith response variable by y_i^m, representing the number of SIDS in month i, $i = 1, \ldots, n$. Then

$$y_i \sim \text{Poisson}(\lambda_j(\mathbf{x}_i^m)) \qquad \text{with probability } p_j(\mathbf{z}_i^m), \tag{7.9}$$

where $\mathbf{X}^m$ and $\mathbf{Z}^m$ are matrices of covariates such that $\mathbf{x}_i^m = [1, x_{1i}^m, x_{2i}^m, \ldots, x_{xgi}^m]'$ and $\mathbf{z}_i^m = [1, z_{1i}^m, z_{2i}^m, \ldots, z_{zgi}^m]'$ with $\mathbf{X}^m$ having a total of xg covariates, and $\mathbf{Z}^m$ zg covariates. The probability density function is then

$$P(y_i^m = q|\mathbf{x}_i^m, \mathbf{z}_i^m) = \sum_{j=1}^{k} \frac{p_j(\mathbf{z}_i^m) \exp(-\lambda_j(\mathbf{x}_i^m))\lambda_j(\mathbf{x}_i^m)^q}{q!} \qquad q = 0, 1, 2, \ldots \tag{7.10}$$

where k represents the number of components in the model. Using logit and log-linear links to model $p_j(\mathbf{z}_i^m)$ and $\lambda_j(\mathbf{x}_i^m)$ respectively gives:

$$\begin{aligned}
\text{logit}(p_j(\mathbf{z}_i^m)) &= \log\left[\frac{p_j(\mathbf{z}_i^m)}{1 - p_j(\mathbf{z}_i^m)}\right] = \mathbf{z}_{ji}^m \boldsymbol{\alpha}_j, \qquad j = 1, \ldots, k-1, \quad i = 1, \ldots, n \\
p_k(\mathbf{z}_i^m) &= 1 - \sum_{j=1}^{k-1} p_j(\mathbf{z}_i) \\
\log(\lambda_j(\mathbf{x}_i^m)) &= \mathbf{x}_{ji}^m \boldsymbol{\beta}_j, \qquad j = 1, \ldots, k \quad i = 1, \ldots, n
\end{aligned} \tag{7.11}$$

for unknown parameters $\boldsymbol{\beta}_j$ and $\boldsymbol{\alpha}_j$. Note that the covariates in the mixing proportions ($\mathbf{Z}^m$) are not restricted to being the same as those in the Poisson rates ($\mathbf{X}^m$). Also, the regression coefficients ($\boldsymbol{\beta}_j$) may vary between components, or be zero for one or more covariates; that is $\boldsymbol{\beta}_{ji} = 0$ for some j ($j = 1, \ldots, k$, $i = 1, \ldots, xg$) implying that the predictor corresponding to $\boldsymbol{\beta}_i$ does not have an impact in the jth component.

The mean and variance of the FM model in equation 7.10 depend on both $\mathbf{X}^m$ and $\mathbf{Z}^m$:

$$\begin{aligned}
E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] &= \sum_{j=1}^{k} p_j(\mathbf{z}_i^m)\lambda_j(\mathbf{x}_i^m) \\
Var[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] &= E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] + \Big[\sum_{j=1}^{k} p_j(\mathbf{z}_i^m)\lambda_j(\mathbf{x}_i^m)^2 - \big(\sum_{j=1}^{k} p_j(\mathbf{z}_i^m)\lambda_j(\mathbf{x}_i^m)\big)^2\Big] \\
&= E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m]\Big[1 + \lambda_j(\mathbf{x}_i^m) - E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m]\Big].
\end{aligned} \tag{7.12}$$

The inclusion of the second term in the variance means that $Var[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] > E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m]$, therefore implying that overdispersion (with respect to the Poisson distribution) is included in the underlying model formulation.

Fitting the Finite Mixture Model

Estimation of model parameters for the FM models is achieved following Wang's approach (1998), where maximum likelihood estimates (MLEs) are obtained using a combination of the EM and quasi-Newton algorithms. The maximum likelihood formulation is as follows.

Define $\mathcal{Z}_{ji}$ as an unobserved indicator variable representing component membership. For example, $\mathcal{Z}_i = [0, 1, 0, \ldots, 0]'$ indicates that the ith observation belongs in the second mixture component. The log-likelihood for the complete data, where the $\mathcal{Z}_{ji}$s are considered as missing values, is:

$$L_{FM}(\alpha, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{k} \mathcal{Z}_{ji} \log[p_j(\mathbf{z}_i^m) \text{Poisson}(\lambda_j(\mathbf{x}_i^m)]. \tag{7.13}$$

Wang's algorithm (1996) and Fortran code† are used to find MLEs. Parameter estimates from the standard Poisson regression models (Chapter 6) are used as starting values for the EM algorithm. Asymptotic normality of the MLEs is utilised to construct approximate standard errors and corresponding p-values.

7.1.3 Zero-inflated Poisson Model

Zero-inflated Poisson (ZIP) models provide a third way to model count data with excess zeros. In this case, the response variable is modelled as a mixture of a Bernoulli distribution and a Poisson distribution. Although ZIP models without covariates have a long history (for example, Cohen (1963), or Johnson & Kotz (1969)), it is only recently Lambert (1992) provided the general form of ZIP regression models incorporating covariates. Lambert presented an application of defects in manufacturing where it was suggested that "one interpretation [of the excess zero distribution] is that slight, unobserved changes in the environment cause the process to move randomly back and forth between a perfect state in which defects are extremely rare and an imperfect state in which defects are possible but not inevitable." The existence of such a perfect state increases the number of zero counts in the data.

ZIP regression models have also been utilised in other applications, including assessing the effect of foetal growth and postnatal somatic growth on fine motor exercise (Cheung, 2002), modelling young driver motor vehicle crashes (Lee et al., 2002), and examining a decayed, missing and filled teeth index in dental epidemiology (Böhning et al., 1999).

In ZIP regression, the number of zeros are inflated by combining a mass at zero with a Poisson distribution. The independent counts are again denoted y_i^m, $i = 1, \ldots, n$, then

$$\begin{aligned} y_i^m &= 0 && \text{with probability } 1 - p(\mathbf{z}_i^m) \\ y_i^m &\sim \text{Poisson}(\lambda(\mathbf{x}_i^m)) && \text{with probability } p(\mathbf{z}_i^m), \end{aligned} \tag{7.14}$$

where $\mathbf{X}^m$ and $\mathbf{Z}^m$ again represent covariate matrices. The probability density function for y_i^m, assuming the ZIP model formulation is then

$$\begin{aligned} P(y_i^m = 0 | \mathbf{x}_i^m, \mathbf{z}_i^m) &= 1 - p(\mathbf{z}_i^m) + p(\mathbf{z}_i^m) \exp(-\lambda(\mathbf{x}_i^m)) \\ P(y_i^m = q | \mathbf{x}_i^m, \mathbf{z}_i^m) &= \frac{p(\mathbf{z}_i^m) \exp(-\lambda(\mathbf{x}_i^m)) \lambda(\mathbf{x}_i^m)^q}{q!} \qquad q = 1, 2, \ldots. \end{aligned} \tag{7.15}$$

Here $(1 - p(\mathbf{z}_i^m))$ is the probability of the so-called perfect state (Lambert, 1992) and $p(\mathbf{z}_i^m)$ is the probability that the number of events follows a Poisson distribution. Conditional on the imperfect state, $\lambda(\mathbf{x}_i^m)$ is the mean number of events occurring. A logit and log-linear link are used to model $p(\mathbf{z}_i^m)$ and $\lambda(\mathbf{x}_i^m)$ respectively, with covariates $\mathbf{X}^m$ and $\mathbf{Z}^m$ potentially differing, as with the FM model.

The (conditional) mean and variance (Cameron & Trivedi, 1998) for the ZIP model defined in equation 7.15 is:

†Available at http://coe.ubc.ca/users/marty/research.html

$$\begin{aligned} E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] &= p(\mathbf{z}_i^m)\lambda(\mathbf{x}_i^m) \\ Var[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] &= E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] + p(\mathbf{z}_i^m)(1 - p(\mathbf{z}_i^m))\lambda(\mathbf{x}_i^m) \\ &= E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m]\Big[1 + \lambda(\mathbf{x}_i^m) - E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m]\Big]. \end{aligned} \tag{7.16}$$

It can be seen clearly that $E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] < Var[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m]$, again forcing overdispersion with respect to the Poisson distribution into the underlying model formulation.

Following the derivations of equations 7.14 to 7.16, the ZIP model can be viewed as a special case of a two component FM model, with no covariates in the mixing probabilities (Wang et al., 1998). One component is taken as a degenerate distribution, with mass at $y_i^m = 0$, while the other component is a Poisson regression model. The ZIP model is more restrictive than the general FM model in that it only allows mixing with respect to zeros.

Fitting the Zero-inflated Poisson Model

Lambert (1992) developed an EM algorithm which maximises the log-likelihood for ZIP regression. This involves fitting two general linear models (GLMs): a logistic and a Poisson regression model. The log-likelihood is:

$$\begin{aligned} L_{ZIP}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = &\sum_{y_i^m=0} \log(\exp(\mathbf{z}_i^m\boldsymbol{\alpha}) + \exp(-\exp(\mathbf{x}_i^m\boldsymbol{\beta}))) + \sum_{y_i^m>0} (y_i^m\mathbf{x}_i^m\boldsymbol{\beta} - \exp(\mathbf{x}_i^m\boldsymbol{\beta})) \\ &- \sum_{i=1}^{n} \log(1 + \exp(\mathbf{z}_i^m\boldsymbol{\alpha})) - \sum_{y_i^m>0} \log(y_i^m!). \end{aligned} \tag{7.17}$$

Implementation of the EM algorithm was achieved using the 'ZIP' function in STATA (*Intercooled STATA, Version 6.0; Stata Corporation*). Standard errors and corresponding significance levels for the parameter estimates are calculated using asymptotic normal theory.

7.1.4 Hurdle Model

The hurdle model is a two part model. The first part consists of a binomial probability model, which determines whether a zero or non-zero outcome occurs. A truncated count data distribution, which describes the positive outcomes, is modelled as the second part. The idea behind this formulation is that given an event has occurred, that is, the "hurdle has been crossed", the conditional distribution of this event is controlled by a truncated-at-zero distribution (Cameron & Trivedi, 1998). Hurdle models have been used in ecological modelling, primarily in applications of species abundance (Lindenmayer et al., 2003; Welsh et al., 1996) and recreational fishing catch data (O'Neill & Faddy, 2003).

The hurdle specification is defined as follows in the Poisson case: the counts y_i^m, $i = 1, \ldots, n$ are independent, and

$$\begin{aligned} &y_i^m = 0 && \text{with probability } 1 - p(\mathbf{z}_i^m) \\ &y_i^m \sim \text{truncated Poisson}(\lambda(\mathbf{x}_i^m)) && \text{with probability } p(\mathbf{z}_i^m), \end{aligned} \tag{7.18}$$

where $\mathbf{X}^m$ and $\mathbf{Z}^m$ are matrices of covariates. The hurdle model is then given by

$$
\begin{aligned}
P(y_i^m = 0|\mathbf{z}_i^m) &= 1 - p(\mathbf{z}_i^m) \\
P(y_i^m = q|\mathbf{x}_i^m, \mathbf{z}_i^m) &= \frac{p(\mathbf{z}_i^m)\exp(-\lambda(\mathbf{x}_i^m))\lambda(\mathbf{x}_i^m)^q}{q!(1-\exp(-\lambda(\mathbf{x}_i^m)))} \qquad q = 1, 2, \ldots .
\end{aligned} \tag{7.19}
$$

This formulation, presented by Welsh et al. (1996), increases the probability of the zero outcome and scales the remaining probabilities so that they add to one. Logit and log-linear functionals model $p(\mathbf{z}_i^m)$ and $\lambda(\mathbf{x}_i^m)$ respectively. Correspondingly, the covariates $\mathbf{X}^m$ are not restricted to being the same as the covariates $\mathbf{Z}^m$.

The form of the conditional mean and variance for the hurdle model is determined by a combination of the probability of crossing the hurdle ($p(\mathbf{z}_i^m)$) and the mean and variance of the truncated Poisson distribution (Cameron & Trivedi, 1998). That is:

$$
\begin{aligned}
E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] &= \frac{p(\mathbf{z}_i^m)\lambda(\mathbf{x}_i^m)}{1-\exp(-\lambda(\mathbf{x}_i^m))} \\
Var[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m] &= \frac{\lambda(\mathbf{x}_i^m)}{1-\exp(-\lambda(\mathbf{x}_i^m))}\Big[1-\exp(-\lambda(\mathbf{x}_i^m))E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m]\Big].
\end{aligned} \tag{7.20}
$$

Again, the mean and variance differ, implying that overdispersion is induced into the model, with respect to the Poisson distribution.

The hurdle model formulation is very similar to the ZIP model, with both models essentially combining binomial probabilities with Poisson distributions. The hurdle model keeps the zero-class distinct from the non-zeros by modelling the non-zero y_i^m's with a truncated Poisson distribution. This differs from the ZIP model where zeros can occur both in the 'perfect state' (with probability $1 - p(\mathbf{z}_i^m)$) and the Poisson distribution (with probability $p(\mathbf{z}_i^m)\exp(-\lambda(\mathbf{x}_i^m))$). In contrast, the FM model can combine both zeros and positive counts from two or more underlying densities. The FM, ZIP and hurdle model formulations all involve discrete mixture distributions, in comparison to the continuous Poisson-gamma mixture defining the negative binomial model.

Fitting the Hurdle Model

The log-likelihood for the hurdle model is:

$$
\begin{aligned}
L_{Hurdle}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{y_i^m=0} \log\left(\frac{1}{1+\exp(\mathbf{z}_i^m\boldsymbol{\alpha})}\right) + \sum_{y_i^m>0} \log\left(\frac{\exp(\mathbf{z}_i^m\boldsymbol{\alpha})}{1+\exp(\mathbf{z}_i^m\boldsymbol{\alpha})}\right) \\
&\quad + \sum_{y_i^m>0}\Big(y_i^m\mathbf{x}_i^m\boldsymbol{\beta} - \exp(\mathbf{x}_i^m\boldsymbol{\beta}) - \log(1-\exp(-\exp(\mathbf{x}_i^m\boldsymbol{\beta}))) - \log(y_i^m!)\Big) \\
&= L_{Hurdle}(\boldsymbol{\alpha}) + L_{Hurdle}(\boldsymbol{\beta})
\end{aligned} \tag{7.21}
$$

A MLE procedure is used to model the SIDS data, as was the case with the FM and ZIP methods. The two log-likelihood components of the hurdle model ($L_{Hurdle}(\boldsymbol{\alpha})$ and $L_{Hurdle}(\boldsymbol{\beta})$) may be fitted separately as (7.21) is the sum of two distinct likelihoods. The first component, $L_{Hurdle}(\boldsymbol{\alpha})$, is a likelihood based on a logistic model and, as such, is fitted using the LOGISTIC procedure in SAS (*SAS (R) Proprietary Software Release (8.01.01)*). The second component, $L_{Hurdle}(\boldsymbol{\beta})$, is fitted by a specifically constructed Matlab macro which uses the Nelder-Mead algorithm to maximise $L_{Hurdle}(\boldsymbol{\beta})$ (*Matlab, Version 6.1.0.1989a, Release 12.1; The MathWorks, Inc.*). Welsh et al. (1996) present details on obtaining the standard errors and corresponding normal-theory p-values for the estimated model parameters.

7.1.5 Extended Poisson Process Model

Faddy (1997) presented a method which analyses count data in terms of an extended Poisson process (Cox & Miller, 1977). These extended Poisson process (EPP) models can incorporate both overdispersion, or underdispersion, with respect to the Poisson distribution, and include both Poisson and negative binomial models. Applications of the EPP model have included animal abundance (Faddy, 1996), lesion counts (Faddy, 1997), and epilepsy treatments (Toscas & Faddy, 2003). Details for the EPP model formulation given below follow the notation of Faddy (1997).

A simple Poisson process is defined as a series of 'events' occurring over time t, where the probability of an event occurring in the interval $(t, t+\delta t)$ is $\lambda\delta t + o(\delta t)$, which is independent of the occurrence of events up to time t (Cox & Miller, 1977). Any discrete probability distribution $\{\pi_0, \pi_1, \pi_2, \dots\}$ can be described in terms of an extended Poisson process, $\{Y(t) : t \geqslant 0\}$ (a Markov birth process) with $Y(0) = 0$, where the transition probabilities are given by

$$P\{Y(t+\delta t) = i+1 | Y(t) = i\} = \lambda_i \delta t + o(\delta t), \qquad t \geqslant 0. \tag{7.22}$$

That is, there exists a sequence $\lambda_0, \lambda_1, \lambda_2, \dots$ $(\lambda_i \geqslant 0)$ such that

$$P\{Y(t) = i\} = \pi_i \tag{7.23}$$

for $i = 0, 1, 2, \dots$ and a fixed time t.

If this sequence is increasing, that is $\lambda_0 < \lambda_1 < \lambda_2 < \dots$, then $Y(t)$ is overdispersed relative to the Poisson distribution; that is $Var[Y(t)] > E[Y(t)]$. In contrast, if the sequence is decreasing, that is $\lambda_0 > \lambda_1 > \lambda_2 > \dots$, then $Y(t)$ is underdispersed relative to the Poisson distribution.

The probabilities $p_i(t) = P(Y(t) = i)$, $i = 0, 1, 2, \dots$, can be defined in terms of the Kolomogorov forward differential equations (Cox & Miller, 1977):

$$\begin{aligned} \frac{dp_0(t)}{dt} &= -\lambda_0 p_0(t), && \text{with } p_0(0) = 1 \\ \frac{dp_i(t)}{dt} &= -\lambda_i p_i(t) + \lambda_{i-1} p_{i-1}(t), && \text{with } p_i(0) = 0, \quad i \geqslant 1. \end{aligned} \tag{7.24}$$

The solution of these differential equations is

$$[p_0(t), p_1(t), p_2(t), \dots, p_i(t)] = [1, 0, 0, \dots, 0] \exp_m\{\mathbf{Q}t\}, \tag{7.25}$$

where the matrix $\mathbf{Q}$ takes the form

$$\mathbf{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots & 0 & 0 \\ 0 & -\lambda_1 & \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -\lambda_2 & \lambda_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\lambda_{i-1} & \lambda_{i-1} \\ 0 & 0 & 0 & 0 & \dots & 0 & -\lambda_i \end{bmatrix}$$

and $\exp_m$ represents the matrix-exponential function (Golub & Loan, 1996). Equation 7.25 is then determined numerically using the Nelder-Mead algorithm via the '*fminsearch*' function in Matlab (*Matlab, Version 6.1.0.1989a, Release 12.1; The MathWorks, Inc.*), where, without loss of generality, t can be taken as one. In the application to modelling monthly SIDS counts, this is achieved by matching the probabilities

$p_0(1), p_1(1), p_2(1), \ldots$ (from equation 7.25) to the observed relative frequencies of the monthly SIDS counts, y_i^m.

Faddy (1997) suggests the following functional form to model the λ_is:

$$\lambda_i = a(i+b)^c \qquad a > 0, \quad b > 0, \quad c \leqslant 1. \tag{7.26}$$

This functional form can model overdispersed sequences of the λ_is. Thus distributions, which in terms of their underlying features lie somewhere between the Poisson distribution and the negative binomial distribution, can be modelled. The functional form given in equation 7.26 can also model distributions which are underdispersed relative to the Poisson. The index c plays an important role in the dispersion of the distribution, such that

$$\begin{aligned} c < 0 \quad & Y(t) \text{ is underdispersed} \\ c = 0 \quad & Y(t) \sim \text{Poisson}(\lambda) \\ c = 1 \quad & Y(t) \sim \text{negative binomial.} \end{aligned}$$

The inclusion of the scale parameter b in equation 7.26 ensures that $\lambda_0 > 0$ which enables the process to move from *State 0* to *State 1*. The parameter a is analogous to a gradient coefficient and gives an indication of the 'speed' of the process (Faddy, 1997).

Covariates are incorporated into λ_i through a, such that a is modelled via a log-link function:

$$\log(a) = \mathbf{X}^m \beta. \tag{7.27}$$

Faddy (1997) presents a diffusion approximation for the mean and variance of the process defined in equation 7.22. These are of the following form (where t is taken as one, without loss of generality):

$$\begin{aligned} E[Y(1)|\mathbf{X}^m] &= b\left[(1+r)^{\frac{1}{1-c}} - 1\right] \\ Var[Y(1)|\mathbf{X}^m] &= \frac{b}{1-2c}(1+r)^{\frac{1}{1-c}}\left(1-(1+r)^{\frac{2c-1}{1-c}}\right) \end{aligned} \tag{7.28}$$

where $r = \frac{a(1-c)}{b^{1-c}} = \frac{\exp(\mathbf{X}\beta)(1-c)}{b^{1-c}}$. In the application to modelling SIDS counts, $E[Y(1)|\mathbf{X}^m]$ corresponds to the expected probability distribution of SIDS. Standard errors are calculated using asymptotic normal theory (Mood et al., 1974), as with the previous methods.

7.2 Statistics for Model Comparison

In this chapter, model comparison essentially involves two processes. Firstly, within each of the five formulations, the best model is identified. The second process compares the resulting best models from each of the five formulations, to find which model is considered best overall. The four statistics utilised to make these model comparisons are: the Deviance statistic (D, or G, equations 3.10 and 3.11, page 65); Akaike's information criterion (AIC, equation 3.12, page 65); Schwarz's criterion (SC, equation 3.13, page 65); and Pearson's chi-squared statistic (P, equation 3.16, page 66). The statistics are the same model fit statistics utilised in both Chapter 3 and Chapter 6. The first three statistics (D, AIC and SC) are all based on log-likelihoods. Pearson's chi-squared statistic is based on the Pearson residuals, which for the Poisson based mixture models have the following general definition:

$$r_i^m = \frac{y_i^m - \hat{\mu}_i}{\sqrt{\hat{\sigma}_i^2}} \tag{7.29}$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ correspond to the estimated conditional mean and variance of the appropriate model. The Pearson chi-squared statistic, P, is the sum of the squared residuals, as shown in equation 6.11 (page 115).

7.3 Assessing the Fit of the Model

The final model, identified using the statistics described in the previous section, is examined for its appropriateness and adequacy in terms of how closely it describes the monthly incidence of SIDS. This is achieved in two ways: assessing the overall model fit, and examining the residuals. This procedure for assessing the fit of the Poisson mixture models is similar to that presented in Chapter 6 (Section 6.3).

7.3.1 Overall Model Fit

Assessing the overall fit of the final mixture model involves both a formal goodness-of-fit hypothesis test and examining the predictive ability of the model.

Goodness-of-fit Test

The hypothesis

H_0: All aspects of the fitted model are correct

against

H_A: Not all aspects of the fitted model are correct

is tested using Pearson's chi-squared statistic (equation 6.11, page 115) where under the null hypothesis, $P \sim \chi^2_{n-m}$ (m represents the numbers of parameters in the model) (Wang et al., 1996).

Predictive Ability of the Model

As in Chapter 6, the predictive ability of the final mixture model is examined by comparing the predicted probabilities against the observed frequencies. Denote the observed frequencies by $\bar{p}_j$, where $\bar{p}_j$ is the proportion of the observed values $y_i^m = j$ ($j = 0, 1, \ldots$). The predicted probabilities, denoted $\hat{p}_j$, are defined for each model as:

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^{n} P(y_i^m = j), \tag{7.30}$$

where $P(y_i^m = j)$ represents the appropriate probability density function for y_i^m, conditional on $\mathbf{x}_i^m$ and/or $\mathbf{z}_i^m$.

A comparison of $\bar{p}_j$ with $\hat{p}_j$ allows discrepancies between the observed and predicted probability density functions to be easily identified. This may indicate values where the mixture model is inclined to over or under estimate the observed distribution.

7.3.2 Residual Diagnostics

The general form of the Pearson residual, given in equation 7.29, will be used to examine the mixture model residuals for influential or poorly fitted observations. This diagnostic procedure entails examining four plots involving the Pearson residuals in various forms, as described below.

Pearson Residuals Against the Predicted Mean

The initial plot examined is of the residuals (r_i) against the predicted mean, that is, the conditional expected value for the model. This plot is inspected for any outliers, unusual patterns or discrepancies.

Normal Scores Plot

The second plot utilised for residual diagnostics is a normal scores plot of the residuals. The normal scores plot was described in detail in Section 6.3.2 (page 116), and involves comparing the residuals from the fitted model with the predicted residuals under the assumption that they are in fact normally distributed. Deviations from the linear trend indicate lack of model fit over certain values of the dependent variable (McCullagh & Nelder, 1989).

Leverage Matrix

The ith diagonal entry h_{ii}, of the leverage matrix defined in equation 6.14 (page 116) is plotted against its observation number (i). This assesses the effect of the ith observation, where, as described previously, an observation is considered influential if it is larger than $2m/n$ (Cameron & Trivedi, 1998).

Autocorrelation Function

The final plot examined in the residual diagnostics process is the autocorrelation function (ACF) of the residuals. As described in Section 4.3.2 (page 84), due to the underlying time series which is inherent in the data set analysed, autocorrelation may be present in the residuals. Deviations from a purely random series, implying dependence between residuals, are identified by ACF values falling outside the range $\pm 2/\sqrt{n}$, in the plot of the ACF against lags.

7.4 Statistical Modelling Methods

The initial analysis step for each of the five model formulations involves identifying the appropriate baseline model. Seasonality ($Temp^{m}_{meanMA30}$ in Period 1, and $\sin(\frac{2\pi t}{12}) + \cos(\frac{2\pi t}{12})$ in Period 2) and the number of infants at risk of SIDS in any month (NAR) are considered at this stage. Models are constructed using the following model forms:

Baseline model

(1) SIDS $\propto$ intercept + season

(2) SIDS $\propto$ intercept + season + NAR

Climate model

(3) SIDS $\propto$ intercept + season + NAR + climate variables.

A baseline model is formed, selecting the best model from (1) or (2) above. These model forms are also examined with the intercept excluded. The additional contribution of the monthly climate covariates (Model (3) above) is examined initially on an individual basis. Any significant climate variables are then considered for inclusion into a multiple climate model.

Finite Mixture Models

With the finite mixture model, the initial step in the analysis procedure involves identifying the correct number of component Poisson distributions, k, to describe the incidence of SIDS. This is achieved by comparing full baseline models (in the form of (2) above) for various k, and utilising the statistics defined for model comparison in Section 7.2 to identify the appropriate number of components. When $k = 1$, the FM model collapses to the standard Poisson regression model presented in Chapter 6. Therefore, only models

with $k > 1$ were considered. The resulting model will be compared with the Poisson regression after results from modelling are presented.

The second stage in the FM modelling involves identifying the appropriate covariates for both the mixing probabilities $p_j(\mathbf{Z}^m)$, and the component Poisson distributions $\lambda_j(\mathbf{X}^m)$, where there are k such distributions $(j = 1, \ldots, k)$. This step again involves identifying a baseline model, and then examines the significance of the addition of climate based covariates to either $p(\mathbf{Z}^m)$, $\lambda(\mathbf{X}^m)$, or both. Initial values for the EM algorithm discussed in Section 7.1.2 were based on the corresponding coefficient values identified in the Poisson regression analysis

Zero-inflated Poisson and Hurdle Models

The method for modelling the ZIP model involves identifying the appropriate covariate forms for both $p(\mathbf{Z}^m)$ and $\lambda(\mathbf{X}^m)$. Again, baseline models are formed and the significance of additional climate covariates examined. To identify the best form for $p(\mathbf{Z}^m)$, the covariates in $\lambda(\mathbf{X}^m)$ were held fixed and similarly, the covariates in $p(\mathbf{Z}^m)$ were fixed when identifying the form of $\lambda(\mathbf{X}^m)$.

The process involved in modelling the hurdle model is the same as that described for the ZIP model.

Negative Binomial and Extended Poisson Process Models

The NB and EPP models both involve one set of covariates, namely $\mathbf{X}^m$. Therefore the modelling process is straight forward for both these model formulations, with only one corresponding parameter set to estimate. The process described above, with model forms (1), (2) and (3), is followed.

Each step of the modelling process is completed separately for Period 1 (1968—1972) and Period 2 (1973—1989), with climate variables deseasoned if necessary.

7.5 Results

NOTE: Models analogous to those presented in this chapter, but containing both NAR and an intercept are presented in Appendix G.

7.5.1 Period 1 (1968—1972)

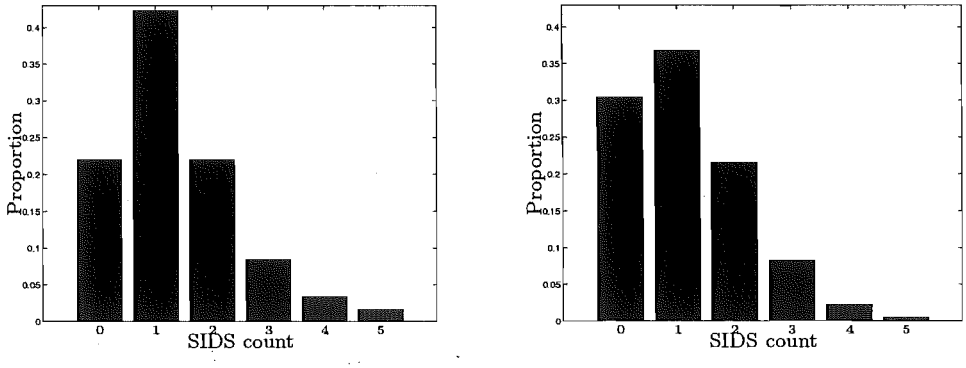
Baseline Models

Table 7.2 presents goodness-of-fit statistics for the saturated baseline FM models, assuming one, two and three components. The three baseline covariates for Period 1 consist of an intercept term, a seasonality term $Temp^m_{meanMA30}$, and NAR, the number of infants at risk of SIDS in any one month. The $k = 1$ component model is the standard Poisson regression model, P1(2) in Table 6.1, (page 120), and is included in Table 7.2 for completeness and ease of comparison between the models. The two component model, which has $m = 9$ parameters, ranks above the three component model across all four goodness-of-fit statistics presented. This $k = 2$ component model also ranks above the $k = 1$ component model for both the deviance (D) and Pearson's chi-squared (P), yet when penalties are added for model complexity, the $k = 1$ component model, with only three parameters, out ranks the two component model. The two component model is utilised to examine the SIDS—climate monthly relationship in Period 1 in terms of the FM model formulation.

Table 7.3 presents the baseline models for Period 1 for the five methods examined in this chapter. With a choice of only three covariates for inclusion in the baseline models, the five models presented necessarily show similar patterns. In terms of the covariates included in the Poisson rate parameter, $\lambda(\mathbf{x}^m)$, the NB, FM, ZIP and EPP best fit baseline models are all of the form $intercept + Temp^m_{meanMA30}$, while the hurdle model also includes NAR. The second Poisson rate component in the FM model is comprised solely of an intercept term. There are differences with respect to the covariates in the mixing probabilities $p(\mathbf{z}^m)$, with

k	m	D	AIC	SC	P
1	3	166.664	172.664	178.896	46.646
2	9	160.208	178.208	196.906	42.442
3	15	164.801	194.801	225.964	46.965

Table 7.2: Saturated baseline Finite Mixture models for determination of the number of components k, Period 1 (m corresponds to the number of parameters in the model).



(a) Observed proportion of SIDS ($\bar{p}$)

(b) Predicted proportion of SIDS ($\hat{p}$)

Figure 7.1: Observed and predicted proportions of SIDS counts, Period 1 (predicted proportions calculated via EPP model).

the FM mixing probabilities defined by an intercept term and NAR. In contrast, both the ZIP and hurdle model contain only the seasonality term, $Temp^{m}_{meanMA30}$.

It must be noted that the standard error corresponding to the parameter value of $Temp^{m}_{meanMA30}$ in the mixing component of the ZIP model is very large, giving a strong indication of lack of model fit (or collinearity). The parameter values in $\lambda(\mathbf{x}^m)$ in both the NB and ZIP models are the same (to three decimal places) as those of the baseline Poisson regression model P1(2) given in Table 6.1 (page 120). Also, the same parameter estimates occur across components in the FM model. The excessively large standard error in the mixing probability of the ZIP model, indicates that the extra zeros inherent in this model's definition are in fact redundant. The dispersion parameter in the NB model ($d = 0.0004$) is very close to zero. The index $c = -4.3$ in the EPP model indicates underdispersion in the SIDS counts, with respect to the Poisson distribution.

The observed distribution of monthly SIDS counts in Period 1, has slightly more ones and less zero counts than what is expected from the theoretical Poisson distribution with $\lambda = 1.339$ (the mean number of SIDS per month in Period 1). With a corresponding observed variance of 1.262, there is also some observed evidence for an underdispersed model. In terms of model formulation, only the EPP model has the ability to model these extra ones, which give rise to the underdispersion in the distribution. Figure 7.1 presents the predicted probability distribution for the baseline EPP model, alongside the observed SIDS distribution for Period 1. The EPP model has not achieved the necessary inflation of the class corresponding to proportion of months with one SIDS death, and has overestimated the zero class, estimating the frequency of zeros at 30% of months, while only 22% of months actually had no SIDS in Period 1.

		NB	FM[i]		ZIP	Hurdle	EPP[j]
			Comp 1	Comp 2			
$\log(\lambda(\mathbf{x}^m))$	Intercept	0.955	0.573	0.955	0.955	24.384	22.67
		(0.0342)	(0.998)	(0.344)	(0.342)	(4.650)	(0.335)
	$Temp^m_{meanMA30}$	-0.059	-0.059		-0.059	-0.086	-0.060
		(0.030)	(0.030)		(0.030)	(0.018)	(0.029)
	NAR					-0.004	
						(0.001)	
$\text{logit}(p(\mathbf{z}^m))$	Intercept		9.709				
			(0.998)				
	$Temp^m_{meanMA30}$				-3.623	-0.094	
					(1647.376)	(0.026)	
	NAR	$d = 0.0004$	0.504				$b = 150.620$
			(0.057)				$c = -4.324$
m		3	5		3	4	4
D		162.950	165.464		166.181	164.518	165.900
AIC		168.950	175.464		172.181	172.518	173.900
SC		175.183	185.851		178.414	180.828	182.210
P		48.800	45.711		48.944	52.944	50.862

Table 7.3: Baseline model parameter estimates (standard errors) for five mixture methods, Period 1.

[i] $k = 2$ chosen as number of components — see Table 7.2.
[j] Modelling $\log(a) = \mathbf{X}^m\beta$.

Climate Models

Table 7.4 presents the best climate model for each of the five methods. In each method, a detailed analysis was completed to identify the best model, across goodness-of-fit statistics. The baseline model forms presented in Table 7.3 were utilised as a guide only.

In terms of the climatic covariates included in the models detailed in Table 7.4, four of the five models contain Dew^{m}_{+diff} in the Poisson rate components. The EPP is the exception, and is also the only model which contains two climate covariates in the rate component, namely $Humid^{dm}_{-diff}$ and $West^{dm}_{+diff}$. Parameter estimates are similar for $Temp^{m}_{meanMA30}$ and Dew^{m}_{+diff} across the Poisson components of the models. Of the three methods which contain the probability component, each contains differing climatic covariates, and the FM, ZIP and hurdle models do not contain the seasonality covariate in this component.

The intercept term in the mixing component of the FM model is extremely large, indicating instability in the model. The index $c = -2.563$ in the EPP model suggests underdispersion in the model with respect to the Poisson distribution, as seen in the baseline form (Table 7.3).

The hurdle model rates best across the likelihood based goodness-of-fit statistics, with the FM model ranked best in terms of Pearson's chi-squared statistic. The ZIP model, in contrast, is consistently the least preferred model, across all four goodness-of-fit statistics. Both the hurdle model (rating best across all four statistics), and the EPP model, the only one of the five methods capable (in terms of the underlying formulation of the model) of incorporating underdispersion into the model, will be considered in more detail. Both overall model fit, and an analysis of the model residuals will be examined.

Assessing the Fit of the Hurdle and EPP Models

Throughout this section of analysis, the hurdle model will be denoted as Model 1, and the EPP model as Model 2. Statistics and variables relating to these models will contain subscripts 'H' or 'EPP' to denote which model the variable is associated with.

Overall Model Fit of the Hurdle and EPP Models

Table 7.4 presents Pearson's chi-squared statistics for both Model 1 and Model 2, where $P_H = 48.9$ ($p = 0.706$, $df = 55$) and $P_{EPP} = 50.9$ ($p = 0.632$, $df = 55$). With respect to both Model 1 and Model 2, the null hypothesis that all aspects of the model are correct cannot be rejected. There is insufficient evidence to imply that the hurdle model, or the EPP model, does not adequately fit the monthly SIDS series.

Figure 7.2 presents the predicted probabilities for both Model 1 and Model 2. Model 1, the hurdle model, shows a large number of months with no SIDS predicted (78%), well in excess of the observed number of 22% (see Figure 7.1 (a)). In contrast, the EPP model does show an inflated number of months where one SIDS death has occurred, but slightly overestimates the class corresponding to two SIDS per month, compared with the observed distribution seen in Figure 7.1. Overall, the EPP model appears to follow the observed distribution reasonably closely, whereas the hurdle model performs very poorly. The hurdle model is therefore no longer considered as a potential candidate for modelling the SIDS—climate relationship in Period 1.

Residual Diagnostics for the EPP Model

The Pearson residuals from Model 2, the EPP model, are plotted against the expected values $E[y_i^m|\mathbf{x}_i^m]$ (as defined in equation 7.28) in Figure 7.3. This plot appears to have bands of residuals corresponding to increasing expected values. Visually, it is difficult to identify poorly fit values from this figure.

The normal score plot of the Pearson residuals from Model 2 is presented in Figure 7.4. The residuals appear linear only over a small range of values, with the deviations from the linear trend largest at the extreme ranges of the residuals. This indicates potential lack of fit with respect to the observed distribution both in months where no SIDS occurred, and months with a few SIDS deaths.

		NB	FM		ZIP	Hurdle	EPP[i]
			Comp 1	Comp 2			
$\log(\lambda(\mathbf{x}^m))$	Intercept		-0.319	0.375		-0.317	6.978
			(0.068)	(0.075)		(0.404)	(0.352)
	$Temp^m_{meanMA30}{}^j$	-0.053	-0.053		-0.045	-0.076	-0.070
		(0.024)	(0.029)		(0.025)	(0.018)	(0.030)
	$Dew^m_{+diff}{}^k$	0.432	0.304		0.617	0.736	
		(0.131)	(0.134)		(0.145)	(0.013)	
	$Humid^{dm}_{-diff}{}^r$						-0.088
							(0.041)
	$West^{dm}_{+diff}{}^s$						-0.308
							(0.149)
$\text{logit}(p(\mathbf{z}^m))$	Intercept		4770.470			-1.337	
			(1128.966)			(0.333)	
	NAR^t		-0.838				
			(1.019)				
	$Dew^m_{mean(std)}{}^u$		72.271				
			(1.036)				
	$Dew^m_{mean(min)}{}^v$				-0.399		
					(0.227)		
	$South^{dm}_{-diff}{}^w$	$d = 0.004$				-1.208	$b = 9.568$
						(0.796)	$c = -2.563$
m		4	7		3	5	6
D		163.468	158.660		189.855	158.411	157.976
AIC		169.468	172.660		195.855	168.411	169.976
SC		175.700	187.203		202.087	178.800	182.441
P		45.910	40.839		49.630	48.720	42.871

Table 7.4: Climate model parameter estimates (standard errors) for the five mixture methods, Period 1 ('d' = deseasoned, m corresponds to the number of parameters in the model).

[i] Modelling $\log(a) = \mathbf{X}^m\beta$.

[j] $Temp^m_{meanMA30}$ = monthly seasonality measure for Period 1.

[k] Dew^m_{+diff} = monthly average when dewpoint was higher on $day0$ than it had been over the past week.

[r] $Humid^{dm}_{-diff}$ = monthly average when humidity was lower on $day0$ than it had been over the past week (deseasoned).

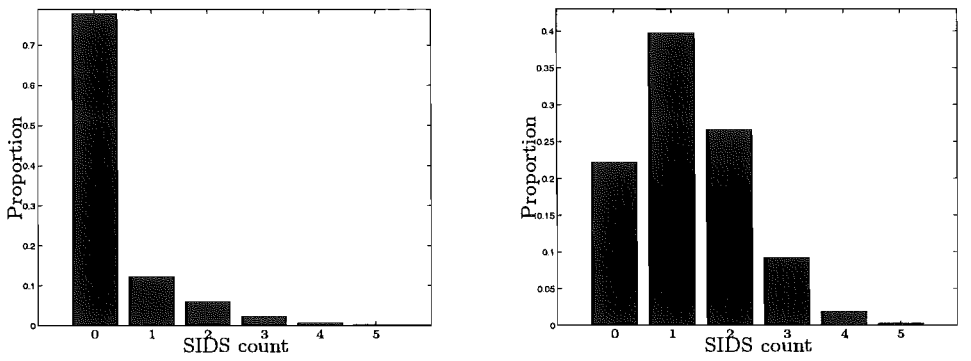
[s] $West^{dm}_{+diff}$ = monthly average when the western component of wind velocity was higher on $day0$ than it had been over the past week (deseasoned).

[t] NAR = number of infants at risk of SIDS.

[u] $Dew^m_{mean(std)}$ = monthly average of the daily standard deviation of dewpoint.

[v] $Dew^m_{mean(min)}$ = monthly average of the daily minimum dewpoint.

[w] $South^{dm}_{-diff}$ = monthly average when the southern component of wind velocity was lower on $day0$ than it had been over the past week (deseasoned).



(a) Hurdle model ($\hat{p}_H$) (b) EPP model ($\hat{p}_{EPP}$)

Figure 7.2: Predicted proportions of SIDS counts for the Hurdle and EPP models, Period 1.

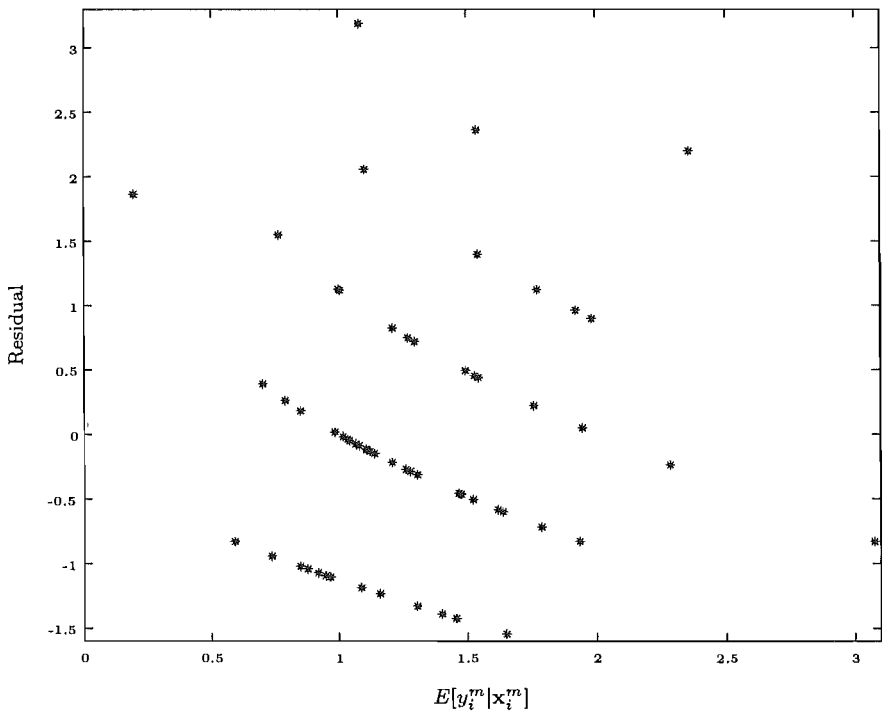


Figure 7.3: Predicted mean (or expected value) against Pearson residuals, Period 1, EPP model.

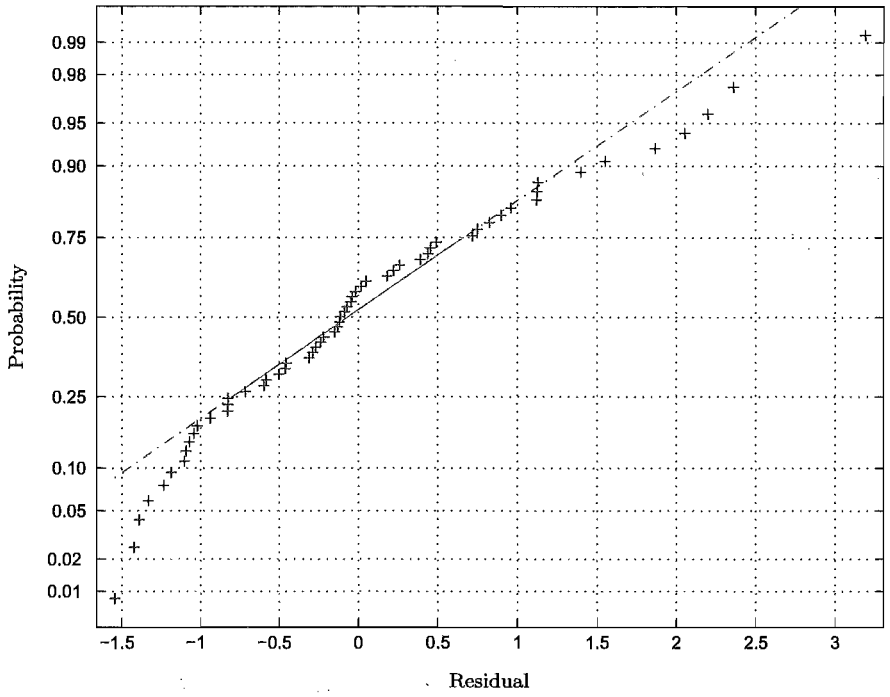


Figure 7.4: Normal scores plot to assess the normality of the Pearson residuals, Period 1, EPP model.

	July 1973
# SIDS	2
Season	winter
$Temp^{m}_{meanMA30}$	6.823
$Humid^{dm}_{-diff}$	-6.625
$West^{dm}_{+diff}$	-0.727

Table 7.5: Covariate values for outlier in EPP model, Period 1.

Figure 7.5 presents the diagonal entries of the leverage matrix (defined in equation 6.14, page 116), which is useful in highlighting influential observations. One outlier falls above the bound $2m/n = 2 \times 6/59 = 0.203$, (shown by – – – – – in Figure 7.5) and is highlighted by ◯. This outlier corresponds to July 1971, and with $h_{43,43} = 0.321$, is well above the bound. As mentioned previously, removal of this point from the dataset is not an option, as this will corrupt the underlying time series inherent in the SIDS data. The covariate pattern corresponding to July 1971 is presented in Table 7.5, and shows that two SIDS occurred in this winter month. No trends are evident that may have resulted in this month being an outlier.

The autocorrelation function for the Pearson residuals from the EPP model is presented in Figure 7.6, for lags of up to 15 months. The ACF shows no evidence of serial correlation in the model, with the series appear purely random. There is a peak at lag$(t - 13)$, but this is not considered extreme.

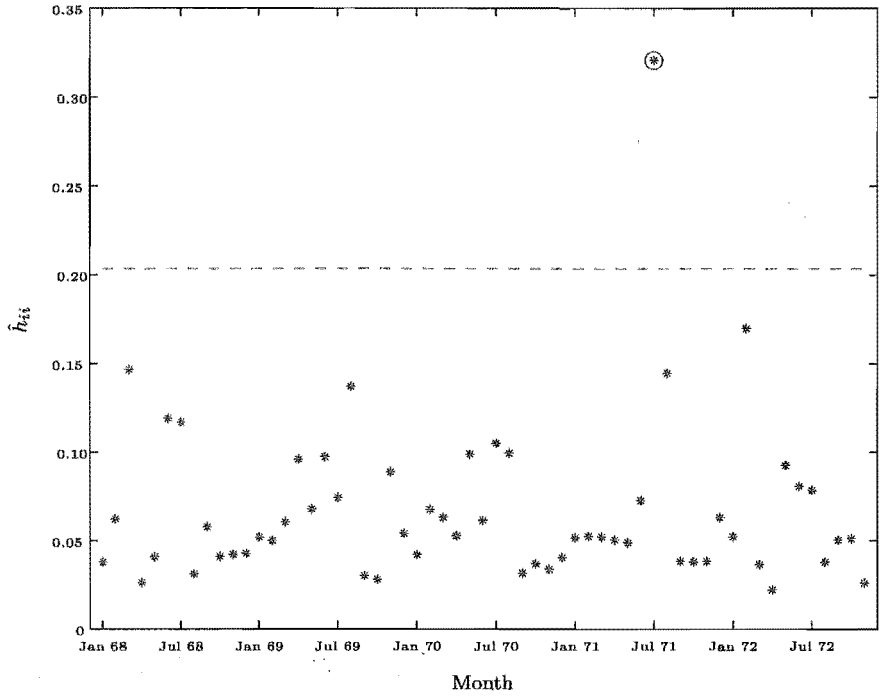


Figure 7.5: Leverage matrix for each observation over time for the EPP model, Period 1.

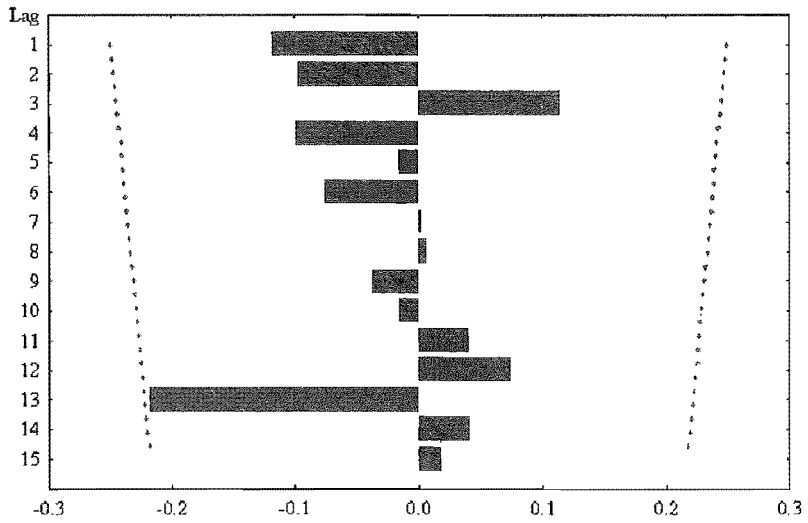


Figure 7.6: Autocorrelation function of the Pearson residuals from the EPP model, Period 1.

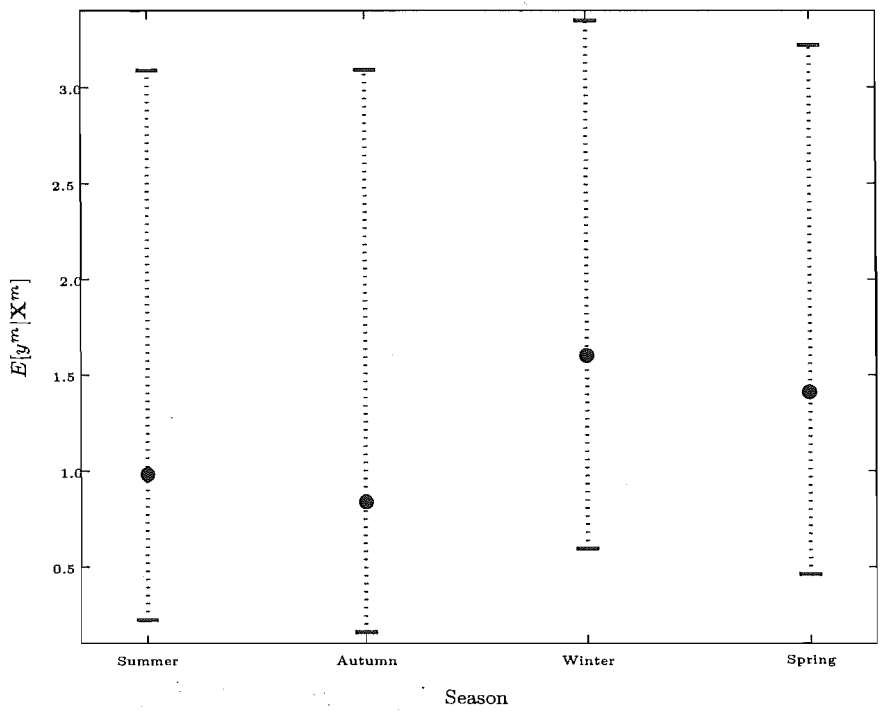


Figure 7.7: Estimated rate of SIDS by season for the EPP model, Period 1 (error bars correspond to 95% confidence intervals of the model coefficients).

Interpretation of the EPP Model

The EPP model is the only one of the five methods which can incorporate underdispersion in its formulation. Examination of the EPP model containing climatic covariates showed that the model adequately described the relationship between SIDS and climate, though fell down somewhat at the extremes of the observed data. This section examines the practical interpretation of the EPP model, with respect to the predicted incidence of SIDS.

The EPP model presented in Table 7.4 contains an intercept term, alongside seasonality defined by $Temp^m_{meanMA30}$, $Humid^{dm}_{-diff}$ and $West^{dm}_{+diff}$. The variable $Humid^{dm}_{-diff}$ describes the relative humidity on $day0$ as compared to what it was over the previous seven days, averaged for the month over days where $Humid_{meanday0}$ was less than the retrospective weekly average. Similarly, the variable $West^{dm}_{+diff}$ describes the strength of western wind velocity on $day0$ as compared to what it was over the previous seven days, averaged for the month over days where $West_{meanday0}$ was more than the retrospective weekly average. Increasing values of both $Humid^{dm}_{-diff}$ and $West^{dm}_{+diff}$ correspond to a decreased risk of SIDS.

Figure 7.7 presents the expected number of SIDS per month for each season, where season averages of the covariates were used to calculate $E[y^m_i|\mathbf{x}^m_i]$. As expected, the expected number of SIDS per month is highest in winter (1.6 SIDS per month), yet the lowest expected number of SIDS per month occurs in autumn (0.8 SIDS per month). With the error bars (corresponding to 95% confidence intervals of the model coefficients), there is little difference between the seasons.

The effect of the two climatic covariates on the expected number of SIDS per month as estimated by the EPP model is shown graphically in Figure 7.8, where the expected number of SIDS per month was calculated holding the remaining covariates fixed at their seasonal averages. The negative relationship between both $Humid^{dm}_{-diff}$ and $West^{dm}_{+diff}$ on the number of SIDS is seen clearly, with this relationship almost linear over

	EPP	Poisson
m	4	2
D	165.9	166.2
AIC	173.9	170.2
SC	182.2	174.3
P	50.9	48.9

Table 7.6: Comparison of goodness-of-fit statistics for EPP model (from Table 7.4, and Poisson model (Model P1(1), Table 6.1, page 120), Period 1 (m corresponds to the number of parameters in the model).

the range of $Humid^{dm}_{-diff}$. Both $West^{dm}_{+diff}$ and $Humid^{dm}_{-diff}$ predict the highest number of SIDS occurring in winter, with $Humid^{dm}_{-diff}$ predicting the lowest number in autumn. In contrast, $West^{dm}_{+diff}$ predicts the lowest number of SIDS occurring in summer. There is little difference between the risk of SIDS in winter or spring, with respect to $West^{dm}_{+diff}$.

Table 7.6 presents goodness-of-fit statistics for the EPP model presented in Table 7.4, alongside those for the Poisson model, P1(1), presented in Table 6.1 (page 120). With half the number of parameters, the Poisson model returned lower goodness-of-fit statistics across three of the four statistics presented. Yet, the Poisson model contained no climatic variables (P1(1) $\propto intercept + Temp^{m}_{meanMA30}$), and failed to predict the large proportion of months where one SIDS occurred (see Figure 6.2, page 121). The EPP model, including two significant climatic covariate effects ($Humid^{dm}_{-diff}$ and $West^{dm}_{+diff}$), was much closer to the observed distribution than that of the Poisson regression model. The additional complexity of the EPP model in comparison to the Poisson (baseline only) model, is considered warranted, as a result of the improved overall model fit (as judged by predicted probability plots (Figure 7.2)), and additional insights provided by $Humid^{dm}_{-diff}$ and $West^{dm}_{+diff}$ into the climate risk profile of SIDS in Period 1.

A comparison of the EPP model with seasonality measured by $Temp^{m}_{meanMA30}$, or by $\sin(\frac{2\pi t}{12}) + \cos(\frac{2\pi t}{12})$, is shown in Table 7.7. These models confirm the choice of $Temp^{m}_{meanMA30}$ as the best measure of seasonality at a monthly level for Period 1; although the deviance for the model containing $\sin(\frac{2\pi t}{12}) + \cos(\frac{2\pi t}{12})$ is marginally smaller than that of the model containing $Temp^{m}_{meanMA30}$, after the addition of penalties for the extra parameter, the model containing $Temp^{m}_{meanMA30}$ is preferred. There is a large difference in Pearson's chi-squared values, 42.8 compared to 63.9 (Table 7.7), with the model containing $Temp^{m}_{meanMA30}$, again being preferred.

7.5.2 Period 2 (1973—1989)

Baseline Models

Table 7.8 presents D, AIC, SC and P values for the saturated baseline FM models, assuming one, two, or three components for the SIDS data. The one component model is simply the standard Poisson regression model and is included in this table for completeness and ease of comparison. The two component model, with twelve parameters, ranks above the three component model (20 parameters) across all three likelihood-based goodness-of-fit statistics, whereas P prefers the more complex three component model. The two component model is selected as the best FM model to examine the SIDS climate relationship. The two Poisson components within the FM model make it directly comparable with both the ZIP and hurdle models, and will be able to highlight differences in climatic risk factors between months of low SIDS and months where there is a high number of SIDS.

The baseline models for Period 2, for the five methods presented in this section of analysis, are given in

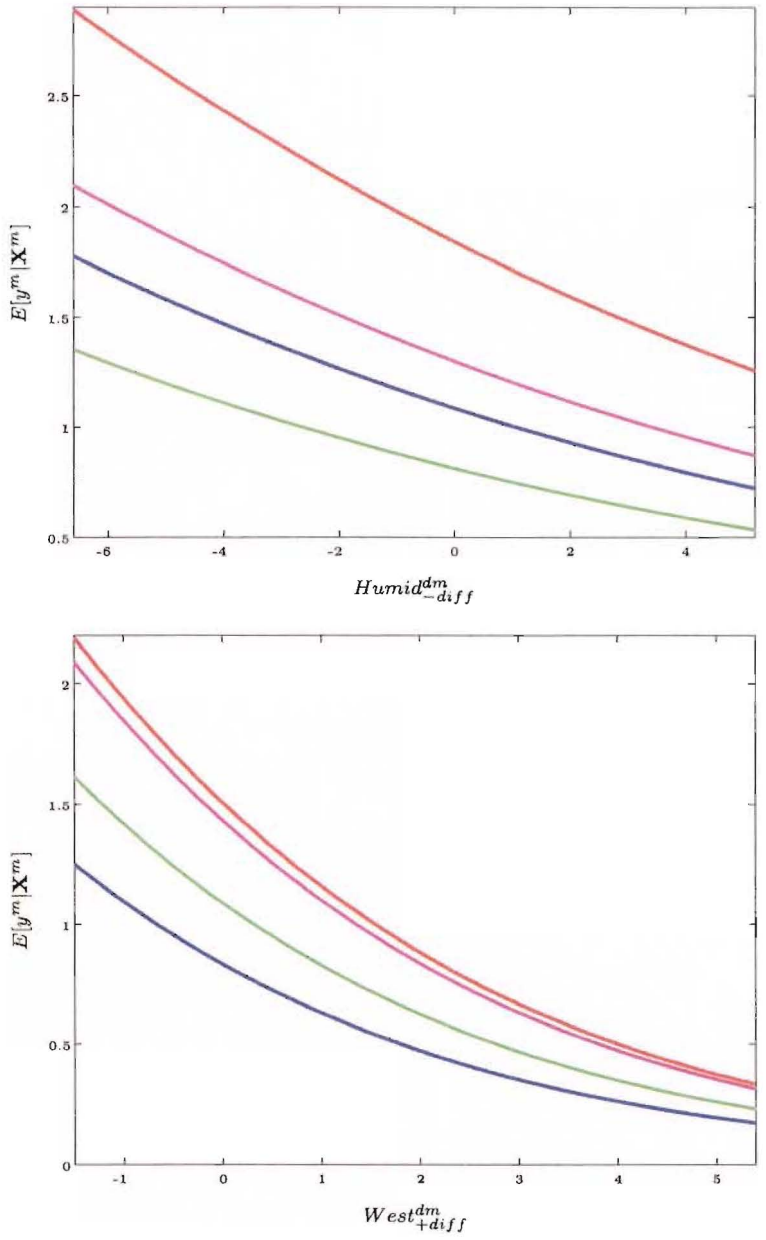


Figure 7.8: EPP model, Period 1, highlighting the effect of $Humid^{dm}_{-diff}$ (top) and $West^{dm}_{+diff}$ (bottom) on the expected number of SIDS per month, by season (— Summer, — Autumn, — Winter, — Spring).

	Model 1	Model 2
Intercept	6.978	2.409
	(0.352)	(0.122)
$Temp^{m}_{meanMA30}$	-0.070	
	(0.030)	
$\sin(\frac{2\pi t}{12})$		-0.387
		(0.175)
$\cos(\frac{2\pi t}{12})$		-0.178
		(0.168)
$Humid^{dm}_{-diff}$	-0.088	-0.068
	(0.041)	(0.044)
$West^{dm}_{+diff}$	-0.308	-0.283
	(0.149)	(0.153)
b	9.568	4.053
c	-2.563	-1.441
m	6	7
D	158.0	156.9
AIC	170.0	170.9
SC	182.4	194.1
P	42.8	63.9

Table 7.7: Comparison of two seasonality measures with the EPP model, Period 1 (m corresponds to the number of parameters in the model).

k	m	D	AIC	SC	P
1	4	718.713	726.713	739.985	196.620
2	12	714.344	738.344	778.162	196.151
3	20	714.920	754.920	821.282	193.946

Table 7.8: Saturated baseline Finite Mixture models for determination of the number of components k, Period 2 (m corresponds to the number of parameters in the model).

Table 7.9. There are many similarities across these baseline models. Both the NB and EPP models have one set of covariates; both models incorporate the seasonality sinusoid, with the NB model also containing an intercept term. In contrast, the EPP baseline model has no intercept term, but instead includes NAR. It has already been noted (Section 6.5.2) that NAR, the number of infants at risk of SIDS, appears to act as a surrogate intercept term.

Within the FM, ZIP and hurdle baseline models, the covariates controlling the proportion of zeros, or mixing proportions ($p(\mathbf{Z}^m)$) differ across all three methods, although no covariate set involves an intercept term. The FM model contains only the sinusoid function in the mixing probability, while, in contrast, the ZIP model's probability of being in the 'perfect state' is comprised solely of NAR. The probability of no SIDS occurring in the hurdle model is defined by both the sinusoid function and NAR.

The Poisson rate parameter in each of the FM, ZIP and hurdle models contains all the covariates considered for inclusion in the baseline models, while the second Poisson component in the FM is defined uniquely by the seasonal sinusoid.

When comparing the goodness-of-fit statistics between the five baseline models presented in Table 7.9, the varying number of parameters contained within the model has a major affect: in terms of SC, which contains the harshest penalty for increased model complexity, models are ranked in a way that directly follows the number of parameters in the model. The NB model which has only four parameters in the model, has the lowest SC value (739.4), and at the other extreme, the FM model, with eight model parameters, has the highest SC value (746.3). The FM baseline model returns the lowest rank sum value, which implies that it rates best over all the goodness-of-fit statistics. The ZIP model, with a rank sum of 11, was the next best model formulation.

Climate Models

The best climatic models for each of the five methods are presented in Table 7.10. In each method, a comprehensive search was undertaken to identify the best model, utilising the baseline model forms presented in Table 7.9 as a guide only, and not forcing this baseline form on the models. Goodness-of-fit statistics were utilised in deciding the best models.

All five models are similarly constructed with $WindS^m_{mean(std)}$ the only significant climate variable in the Poisson rate component of the models. This was in addition to a full baseline model, including an intercept term, seasonal sinusoid and NAR, for four of the five models. Only the NB best model had a restricted form of baseline model, containing only seasonality and $WindS^m_{mean(std)}$, while the second component of the FM model was comprised solely of the sinusoid function. The covariate coefficients are comparable across $\lambda(\mathbf{X}^m)$ for all five models.

There are some variations in the covariate structure of $p(\mathbf{Z}^m)$ between the FM, ZIP and hurdle models. The covariates found to best fit the mixing proportions for the FM model were $sin(\frac{2\pi t}{12}) + cos(\frac{2\pi t}{12})$, the seasonal sinusoid. The best covariate combination for both the ZIP and hurdle model forms contained a climate variable in addition to the seasonal sinusoid. With the hurdle, this was again $WindS^m_{mean(std)}$, while the ZIP model contained $Humid^m_{min(min)}$. The three models containing $p(\mathbf{Z}^m)$ returned differing covariate coefficients.

Table 7.10 also presents goodness-of-fit statistics for each of the five methods, enabling comparison directly in terms of model fit, between the methods. The number of parameters in the models, m, ranged from four in the NB model, to more than double this with the FM model having nine model parameters. This difference is reflected directly in the two penalised likelihood statistics, AIC and SC.

Both the FM model and the ZIP model perform well with respect to the goodness-of-fit statistics, rating first and second respectively, in terms of the overall rank sum values. Both these models will be examined

		NB	FM[i]		ZIP	Hurdle	EPP[j]
			Comp 1	Comp 2			
$\log(\lambda(\mathbf{x}^m))$	Intercept	0.796	1.304		1.668	2.107	
		(0.049)	(0.491)		(0.463)	(0.247)	
	$\sin(\frac{2\pi t}{12})$	-0.326	-0.593	-0.176	-0.318	-0.259	-0.317
		(0.067)	(0.122)	(0.100)	(0.067)	(0.035)	(0.069)
	$\cos(\frac{2\pi t}{12})$	-0.442	-0.038	-0.857	-0.442	-0.400	-0.449
		(0.068)	(0.220)	(0.203)	(0.068)	(0.036)	(0.070)
	NAR		-0.0002		-0.0002	-0.0003	-0.0002
			(0.0001)		(0.0001)	(0.0001)	(0.0001)
$\text{logit}(p(\mathbf{z}^m))$	$\sin(\frac{2\pi t}{12})$		0.913			-1.226	
			(0.213)			(0.360)	
	$\cos(\frac{2\pi t}{12})$		-17.567			-1.253	
			(1.910)			(0.361)	
	NAR	$d = 0.008$			0.001	0.0005	$b = 126.931$
					(0.0004)	(0.0001)	$c = 0.334$
m		4	8		5	7	5
D		718.067	703.778		714.308	707.873	714.667
AIC		726.067	719.778		724.308	721.873	724.667
SC		739.339	746.323		740.898	745.100	741.257
P		200.030	189.441		205.580	236.640	217.870
Rank sum		13	8		11	13	15

Table 7.9: Baseline model parameter estimates (standard errors) for the five mixture methods, Period 2.

[i] $k = 2$ chosen as number of components — see Table 7.8.
[j] Modelling $\log(a) = \mathbf{X}^m\beta$.

in terms of model fit and adequacy with respect to the Canterbury monthly SIDS data.

It is interesting to note that the baseline EPP model, with $c = 0.681$, fell between the class of Poisson and NB models, yet in the EPP model containing the climatic covariate $WindS^m_{mean(std)}$, the index c fell to -0.044, indicating (slight) underdispersion with respect to the Poisson distribution. This form of EPP model performed poorly in comparison to the other models, which all bring overdispersion into the model form. The NB model also performed poorly in comparison to the FM, ZIP and hurdle models, possibly giving evidence to what was highlighted by the baseline EPP model form — that the monthly SIDS data is best modelled by a method where the variance falls somewhere between $\lambda(\mathbf{X}^m)$ and $\lambda(\mathbf{X}^m) + (\frac{1}{\theta})\lambda(\mathbf{X}^m)^2$.

Assessing the Fit of the FM and ZIP Models

Throughout this section of analysis, where necessary the FM model will be denoted as Model 3, and the ZIP model as Model 4. Statistics and variables relating to these models will contain subscripts 'FM' or 'ZIP' to denote which model the variable is associated with.

Overall Model Fit of the FM and ZIP Models

Pearson's chi-squared statistic for the FM and ZIP models are presented in Table 7.10. The statistics are $P_{FM} = 183.4$ ($p = 0.71$, $df = 195$) and $P_{ZIP} = 191.8$ ($p = 0.757$, $df = 196$). With respect to both Model 3 and Model 4, the null hypothesis that all aspects of the model are correct cannot be rejected. There is insufficient evidence to show that the FM model, or the ZIP model, does not adequately fit the monthly SIDS data.

The predicted probabilities, $\hat{p}_{FMj}$, for the FM model were calculated using equation 7.30, where $P(y_i^m = j)$ is given in equation 7.10. Similarly, the predicted probabilities, $\hat{p}_{ZIPj}$, corresponding to the ZIP model were calculated using $P(y_i^m = j)$ as defined in equation 7.15. Figure 7.9 presents the probability distribution of SIDS counts for the observed data (Figure 7.9 (a)), the FM model (Figure 7.9 (b)) and the ZIP model (Figure 7.9 (c)). There are differences in the two predicted probability distributions, including the FM model predicting the largest proportion corresponding to months where one SIDS death occurs ($\hat{p}_{FM1} = 26\%$). In contrast, the largest predicted probability class for the ZIP model is two SIDS per month, where $\hat{p}_{ZIP2} = 21\%$. There is also differences in the predicted proportion of months where no SIDS deaths occurred, with the FM model predicting 22% compared to the 14% predicted by the ZIP model.

The predicted probabilities from the ZIP model appear to follow the distribution of observed counts more closely than the FM model. The ZIP model underestimates the number of months with two SIDS deaths, and overestimates the number of months with four SIDS deaths. The FM model underestimates the number of months corresponding to both no deaths and one SIDS death, compared to the observed distribution. The number of months with two or three SIDS deaths is underestimated by the FM model.

Residual Diagnostics

The Pearson residuals for Model 1 are plotted against the expected values $E[y_i^m|\mathbf{x}_i^m, \mathbf{z}_i^m]$ in Figure 7.10, for both the FM and ZIP models. This plot appears similar to the residual plots in Figure 6.9 (page 129), for the standard Poisson regression model. Both sets of residuals appear to have bands corresponding to increasing mean values. It is difficult visually to identify poorly fit values from this form of residual plot.

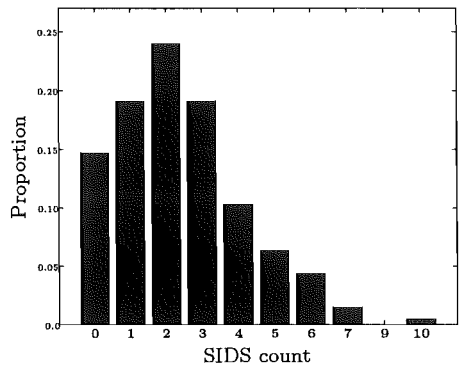
Figure 7.11 presents the normal score plot for the Pearson residuals, for the FM and ZIP models. Both models show the same residual trends. The relationship appears linear for residuals in the range -1 to 1.5, but deviates from the line of normality at the extremes of the residual range. This is especially obvious with residuals at the lower end of the scale (less than -1). The proportion of months where there are either no SIDS, or large numbers of SIDS, is not successfully estimated by either the FM or ZIP models.

The leverage matrix, defined in equation 6.14 (page 116), is useful in highlighting influential observations. Figure 7.12 presents the diagonal entries of the leverage matrix (h_{ii}) plotted over time, for the FM and ZIP

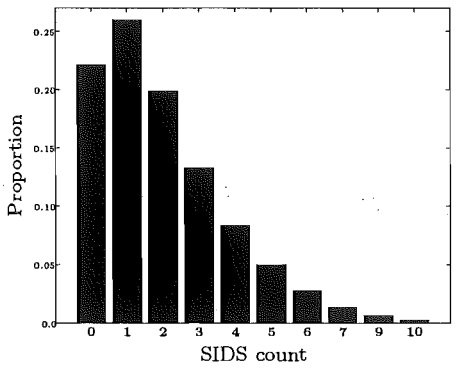
		NB	FM[i]		ZIP	Hurdle	EPPM[j]
			Comp 1	Comp 2			(log(a))
log($\lambda(\mathbf{x}^m)$)	Intercept		0.768		1.178	1.516	1.082
			(0.528)		(0.533)	(0.262)	(0.222)
	$\sin(\frac{2\pi t}{12})$	-0.328	-0.183	-0.577	-0.237	-0.261	-0.322
		(0.067)	(0.095)	(0.115)	(0.072)	(0.033)	(0.028)
	$\cos(\frac{2\pi t}{12})$	-0.444	-0.864	-0.027	-0.409	-0.405	-0.457
		(0.068)	(0.167)	(0.181)	(0.069)	(0.033)	(0.028)
	NAR^r		-0.0002		-0.0002	-0.0003	-0.0002
			(0.0001)		(0.0001)	(0.0001)	(0.0001)
	$WindS^m_{mean(std)}{}^s$	0.177	0.149		0.153	0.158	0.161
		(0.011)	(0.070)		(0.074)	(0.035)	(0.031)
logit($p(\mathbf{z}^m)$)	$\sin(\frac{2\pi t}{12})$		1.248		-5.378	-1.205	
			(0.475)		(2.491)	(0.359)	
	$\cos(\frac{2\pi t}{12})$		-4.544		-1.756	-1.264	
			(0.607)		(1.274)	(0.362)	
	$Humid^m_{min(min)}{}^t$				0.126		
					(0.046)		
	$WindS^m_{mean(std)}$	$d = 0.013$				0.511	$b = 1.220$
						(0.068)	$c = -0.044$
m		4	9		8	8	7
D		714.952	699.611		702.677	703.754	709.670
AIC		722.952	717.611		718.678	719.754	723.670
SC		736.224	747.474		745.222	746.299	746.897
P		194.994	183.425		191.800	198.070	189.970
Rank sum		14	8		9	14	15

Table 7.10: Climate model parameter estimates (standard errors) for the five mixture methods, Period 2.

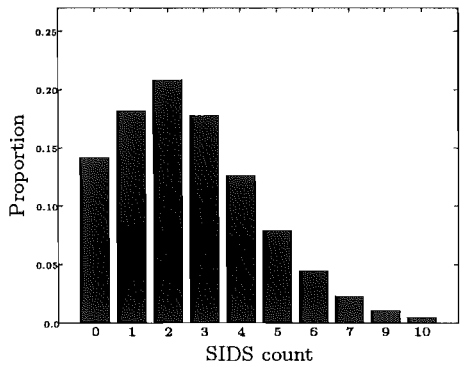
[i] $k = 2$ chosen as number of components — see Table 7.8.
[j] Modelling $log(a) = \mathbf{X}^m\beta$.
[r] NAR = number of infants at risk of SIDS.
[s] $WindS^m_{mean(std)}$ = monthly average of the daily standard deviation of wind speed.
[t] $Humid^m_{min(min)}$ = monthly minimum of the daily minimum humidity.



(a) Observed proportion of SIDS ($\bar{p}$)

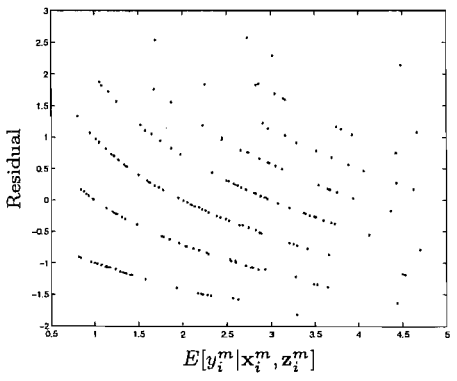


(b) Predicted proportion of SIDS ($\hat{p}_{FM}$)

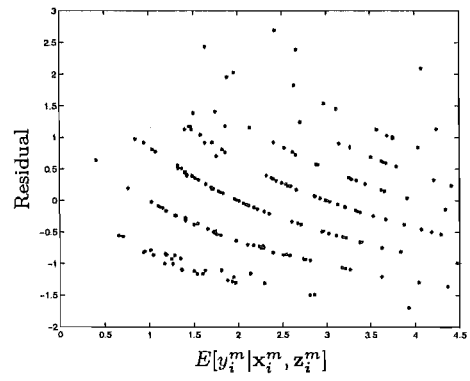


(c) Predicted proportion of SIDS ($\hat{p}_{ZIP}$)

Figure 7.9: Observed and predicted proportions of SIDS counts, Period 2 (predicted proportions calculated via the FM and ZIP models).

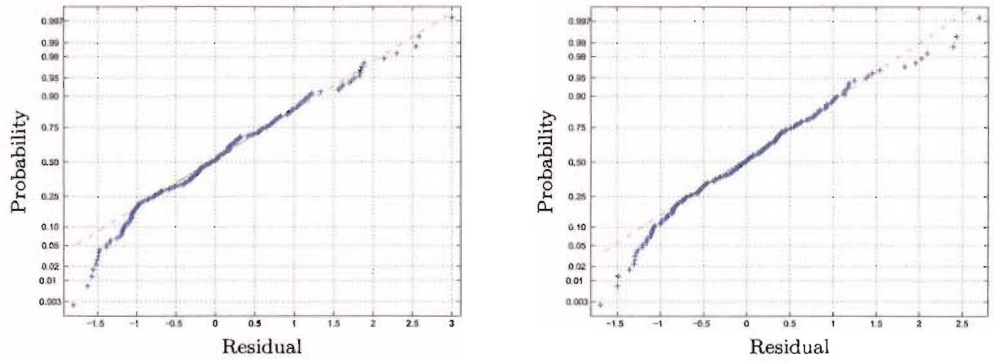


(a) FM model



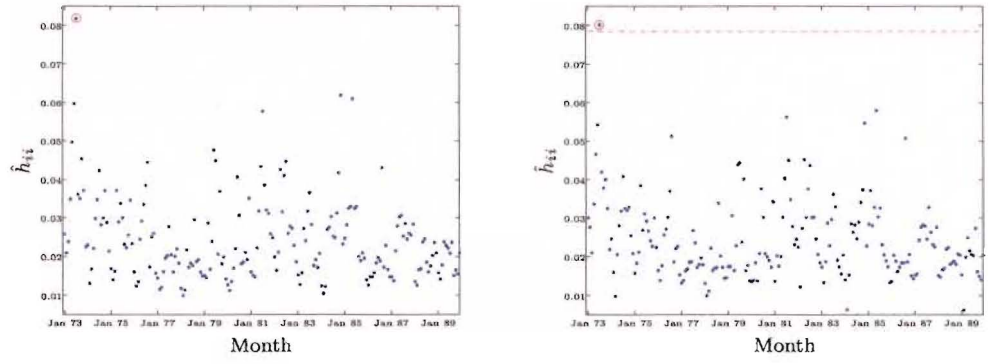
(b) ZIP model

Figure 7.10: Predicted mean (or expected value) against Pearson residuals for both the FM and ZIP models, Period 2.



(a) FM model (b) ZIP model

Figure 7.11: Normal scores plot to assess the normality of the Pearson residuals for both the FM and ZIP models, Period 2.



(a) FM model (b) ZIP model

Figure 7.12: Leverage matrix for each observation over time, for both the FM and ZIP models, Period 2.

models. One potential outlier is highlighted by ○ in each plot. The bound $2m/n = 0.088$ for the FM model, which lies outside the range of $\hat{h}_{ii}$ values pictured in Figure 7.12 (a). The highlighted outlier, corresponding to July 1973, with $\hat{h}_{7,7} = 0.082$ is not considered influential in the FM model formulation.

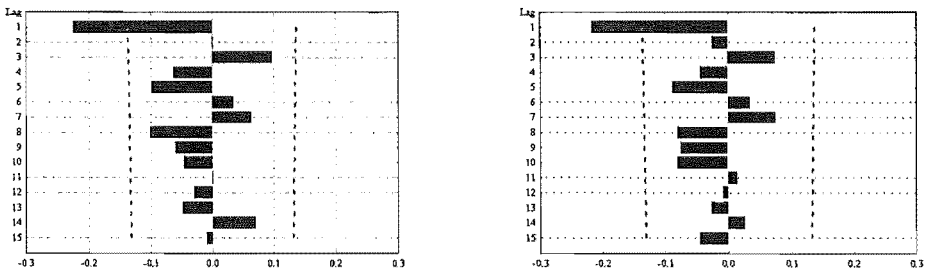
The bound $2m/n = 0.078$ for the ZIP model, and is shown by - - - - - in Figure 7.12 (b). The highlighted outlier again corresponds to July 1973. Covariate values for this month are given in Table 7.11. Four SIDS occurred in this month. No trends are evident within the covariates that may have resulted in this month being an outlying value. Due to the time series nature of the data, this outlier was not removed from the series.

The autocorrelation functions, for the Pearson residuals, for both the FM and ZIP models, are presented in Figure 7.13. The patterns are similar across the models. Both ACFs show a large value at lag$(t-1)$, implying a possible short-term correlation remaining in the residuals. An autoregressive covariate, comprised of lagged SIDS counts was therefore incorporated into both models in the Poisson rate component ($\lambda(\mathbf{z}^m)$).

Table 7.12 presents details of the FM and ZIP models containing the autoregressive covariate, given by

	July 1973
# SIDS	4
Season	winter
$\sin(\frac{2\pi t}{12})$	-0.5
$\cos(\frac{2\pi t}{12})$	-0.866
NAR	5626.5
$WindS^{m}_{mean(std)}$	5.777
$Humid^{m}_{min(min)}$	51.645

Table 7.11: Covariate values for outlier month in the ZIP model, Period 2.



(a) FM model

(b) ZIP model

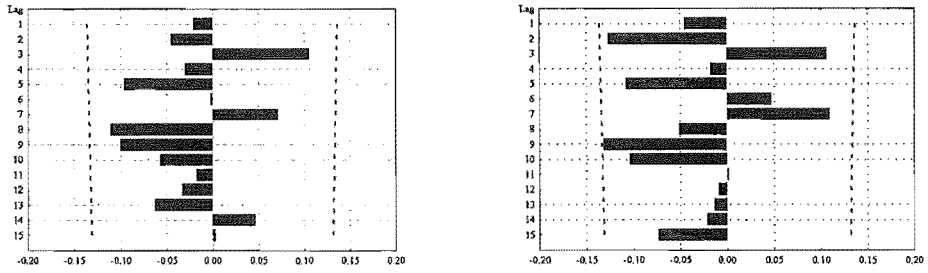
Figure 7.13: Autocorrelation function of the Pearson residuals, for both the FM and ZIP models, Period 2.

		FM		ZIP
		Comp 1	Comp 2	
$\log(\lambda(\mathbf{x}^m))$	Intercept	1.123		1.582
		(0.586)		(0.546)
	$\sin(\frac{2\pi t}{12})$	-0.632	-0.343	-0.341
		(0.176)	(0.099)	(0.078)
	$\cos(\frac{2\pi t}{12})$	-0.025	-1.008	-0.470
		(0.018)	(0.762)	(0.072)
	NAR	-0.0002		-0.0003
		(0.0001)		(0.0001)
	$WindS^m_{mean(std)}$	0.148		0.144
		(0.070)		(0.074)
	y^m_{i-1}	-0.099		-0.094
		(0.033)		(0.029)
$\text{logit}(p(\mathbf{z}^m))$	$\sin(\frac{2\pi t}{12})$	3.760		-5.483
		(0.891)		(2.450)
	$\cos(\frac{2\pi t}{12})$	-14.153		-1.599
		(2.756)		(1.242)
	$Humid^m_{min(min)}$			0.127
				(0.045)
m		10		9
D		688.620		692.122
AIC		708.620		710.122
SC		741.801		739.985
P		173.448		182.950

Table 7.12: Parameter estimates (standard errors) for the FM and ZIP models with autoregressive component y^m_{i-1}, Period 2.

the number of SIDS, lagged by one month (denoted y^m_{i-1}). These models, are both essentially Model + y^m_{i-1}. Inclusion of y^m_{i-1} in the FM model significantly reduced the log-likelihood ($G = 10.991$, $p = 0.001$), which is seen by the lower values of the four goodness-of-fit statistics, as compared with Model 3. The ACF for the residuals from the FM model + y^m_{i-1} is presented in Figure 7.14 (a), and no longer shows any evidence of serial correlation in the time series of residuals. Similarly, including y^m_{i-1} in the ZIP model had a significant impact in terms of lower goodness-of-fit values ($G = 10.555$, $p = 0.001$). Figure 7.14 (b) shows the ACF for the residuals from the ZIP model with the autoregressive component. Like the FM model, this no longer shows any evidence of significance autocorrelation.

Both the FM and ZIP models performed adequately in terms of model fit and residual diagnostics. Both models were similar in their overall performance. The ZIP model appeared to follow the observed probability distribution of SIDS counts more closely than the FM model, yet the FM model rated better than the ZIP model with respect to goodness-of-fit statistics. The model interpretation presented in the next section therefore covers both the FM and ZIP models.



(a) FM model

(b) ZIP model

Figure 7.14: Autocorrelation function of the Pearson residuals, for both the FM and ZIP models with autoregressive term y^m_{i-1}, Period 2.

FM Model Interpretation

The final FM model (presented in Table 7.12) included a single climatic based covariate, namely $WindS^m_{mean(std)}$, alongside variables describing the underlying seasonality in the model ($\sin(\frac{2\pi t}{12})$ and $\cos(\frac{2\pi t}{12})$), the effect of the population at risk of SIDS (NAR), and the underlying autoregressive structure resulting from the time series nature of the data (y^m_{i-1}). The FM model is composed of two Poisson rate components, with a probability term which defines which component is most likely at each month. The covariate structure in both the mixing probability ($p(\mathbf{Z}^m)$) and the second Poisson component ($\lambda_2(\mathbf{X}^m)$) was defined by the seasonality terms, while the first Poisson component ($\lambda_1(\mathbf{X}^m)$) contains an intercept term, alongside $\sin(\frac{2\pi t}{12})$, $\cos(\frac{2\pi t}{12})$, NAR, $WindS^m_{mean(std)}$ and y^m_{i-1}.

Figure 7.15 presents the log of the Poisson rate parameter, $\hat{\lambda}_j(\mathbf{X}^m)$, per season for each component. Mean values of each covariate for each season were used to calculate $\log(\hat{\lambda}_j(\mathbf{X}^m))$. The predicted mean number of SIDS is similar for each component in both autumn and winter: in autumn, the predicted number of SIDS is 1.10 per month and 1.23 per month for component 1 and component 2 respectively, in winter these estimates rise to 2.47 SIDS per month and 2.84 SIDS per month. In contrast, there is a large difference in the estimated mean number of SIDS per month across components in summer and spring: in spring, component 1 predicts 1.35 SIDS per month while component 2 predicts only 0.35 SIDS per month. This difference shifts to a higher predicted SIDS rate per component in spring, but still exists: the predicted number of SIDS is 2.84 and 0.81 per month for component 1 and component 2 respectively.

The predicted probability of belonging in component 1 is essentially zero in both summer and spring, and one in autumn and winter. The estimated number of SIDS per month is therefore strongly dependent on component membership, with summer characterised by component 2, autumn and winter by component 1 and spring again characterised by component 2.

The estimated effect of $WindS^m_{mean(std)}$ in component 1 on SIDS risk is illustrated by season in Figure 7.16, holding the remaining covariates fixed at their seasonal mean value. A one knot increase in the monthly mean of the daily standard deviation in wind speed ($WindS^m_{mean(std)}$) corresponds to the estimated mean number of SIDS per month increasing by 16%. Figure 7.16 shows that the highest risk of SIDS, conditional on belonging in component 1, occurs in spring, whilst the lowest SIDS risk occurs in autumn. The estimated SIDS risk in winter ranges from 2.00 to 3.07 SIDS per month corresponding to $WindS^m_{mean(std)}$ values of 2.90 and 5.80 knots respectively.

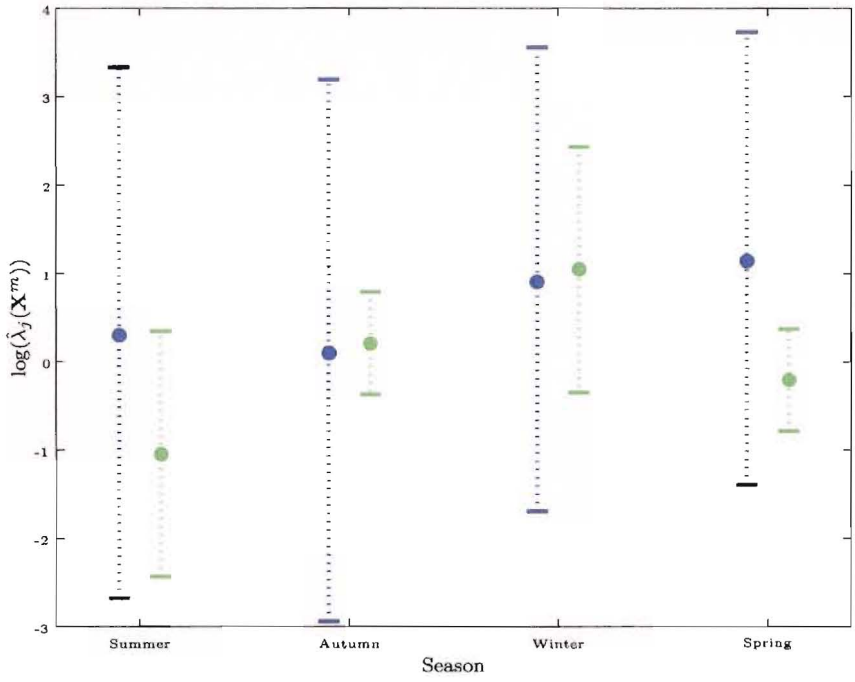


Figure 7.15: Estimated rate of SIDS by season for each component of the FM model, Period 2, with 95% confidence bounds (● component 1, ● component 2).

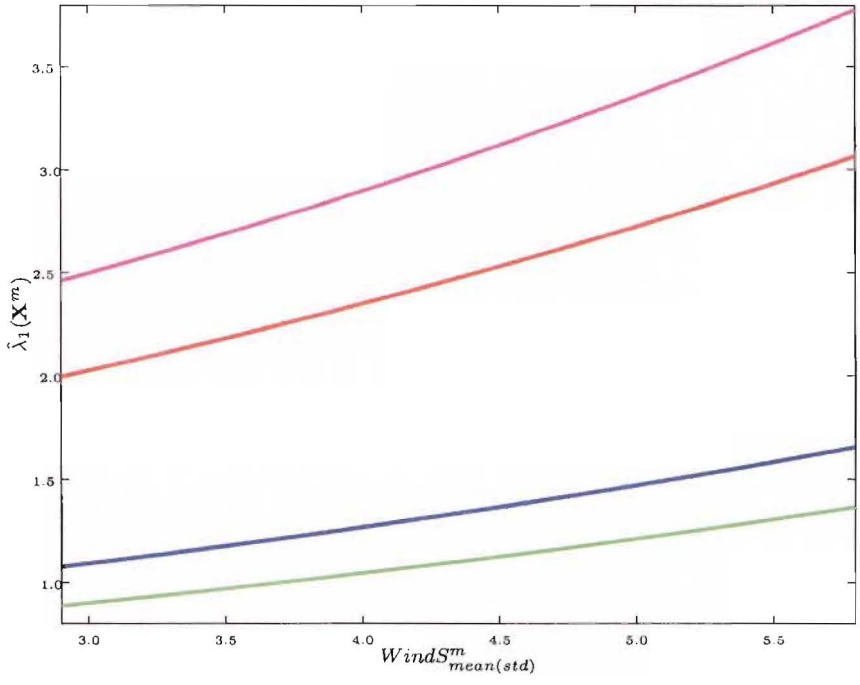


Figure 7.16: FM model, Period 2, highlighting the effect of $WindS^m_{mean(std)}$ on the predicted mean rate of SIDS in component 1, by season (— Summer, — Autumn, — Winter, — Spring).

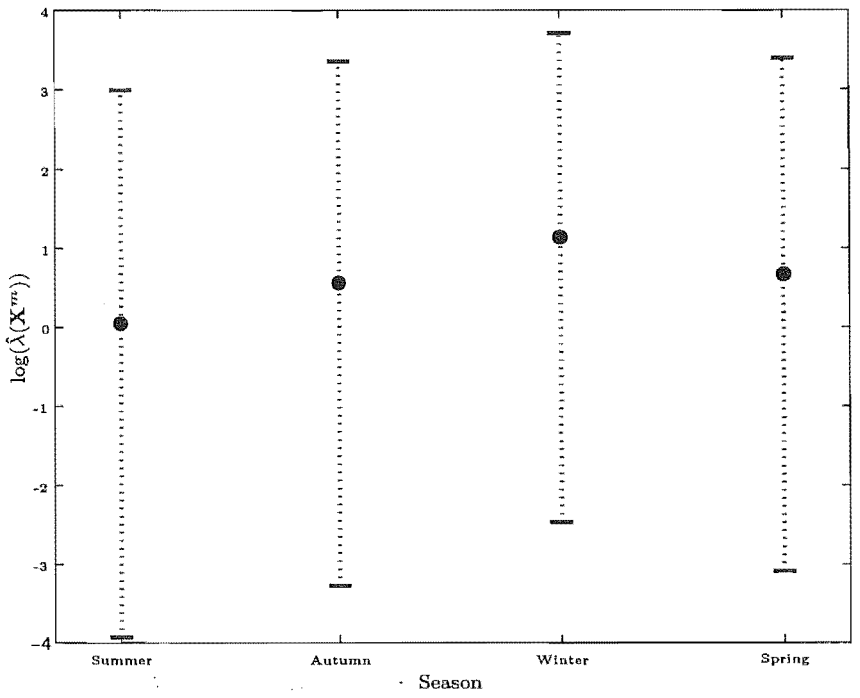


Figure 7.17: Estimated rate of SIDS by season for the ZIP model, Period 2 (with 95% confidence bounds).

ZIP Model Interpretation

The final ZIP model is presented in Table 7.12. This model includes two climatic based covariates: $WindS^m_{mean(std)}$ in the Poisson rate component, and $Humid^m_{min(min)}$ in the probability component. The ZIP model Poisson rate component contains a full baseline form ($intercept$+$\sin(\frac{2\pi t}{12})$+$\cos(\frac{2\pi t}{12})$+NAR), alongside $WindS^m_{mean(std)}$ and the autoregressive covariate y^m_{i-1}. The probability component structure is defined by seasonality terms ($\sin(\frac{2\pi t}{12}) + \cos(\frac{2\pi t}{12})$) alongside humidity.

The log of the estimated Poisson rate component ($\hat{\lambda}(\mathbf{X}^m)$) per season is presented in Figure 7.17 for the ZIP model. Mean seasonal values of the covariates were used to calculate $\log(\hat{\lambda}(\mathbf{X}^m))$. The estimated rate of SIDS ($\hat{\lambda}(\mathbf{X}^m)$) ranges from a low of 1.05 SIDS per month in summer to a peak of 3.12 SIDS per month in winter.

Figure 7.18 presents the estimated effect of $WindS^m_{mean(std)}$ on the Poisson rate component of the ZIP model. Increasing $WindS^m_{mean(std)}$ corresponds to an increase in the risk of SIDS: the lowest estimated risk of 0.85 SIDS per month occurs at 2.9 knots in summer, whereas the highest estimated risk of 3.38 SIDS per month occurs at 5.8 knots in winter. The risk equates to a 15% increase in the SIDS rates for every one knot increase in the monthly average of the daily standard deviation of wind speed. Essentially this implies that the more variable the wind speed is, on average, the higher the estimated SIDS risk. This is the same trend that was evident with respect to wind speed seen in the FM model.

An increase in the minimum monthly humidity ($Humid^m_{min(min)}$) corresponds to a decrease in the probability of belonging to the 'perfect state' (no SIDS deaths occurring). A 10% increase in relative humidity from 45% to 55% decreases the probability of being in the perfect state from 0.013 to 0.004 in summer, and from 0.0008 to 0.0002 in winter.

Table 7.13 presents goodness-of-fit statistics for the FM and ZIP models presented in Table 7.12, alongside

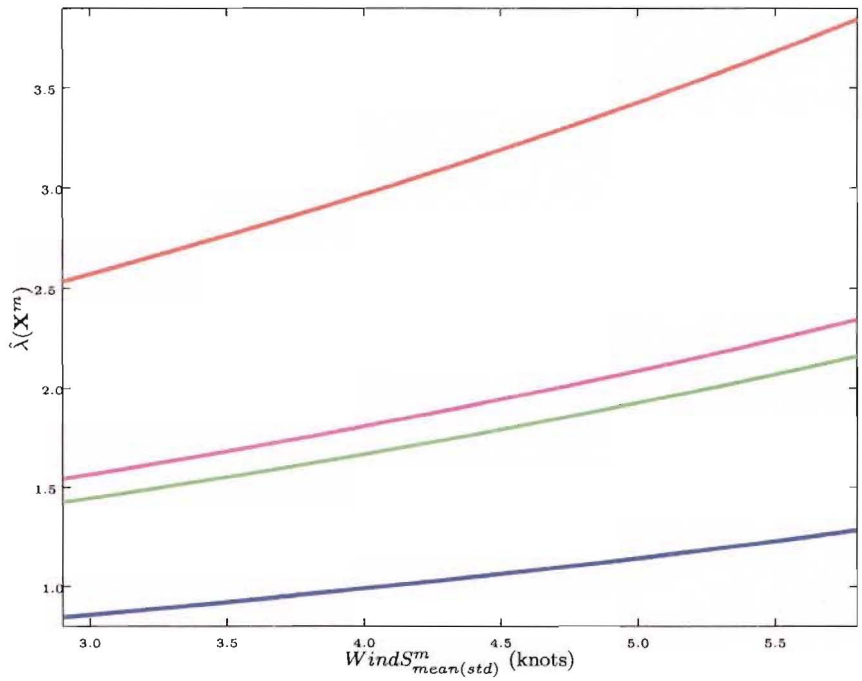


Figure 7.18: ZIP model, Period 2, highlighting the effect of $WindS^m_{mean(std)}$ on the predicted mean rate of SIDS, by season (— Summer, — Autumn, — Winter, — Spring).

	FM	ZIP	Poisson
m	10	9	7
D	688.6	692.1	701.9
AIC	708.6	710.1	715.9
SC	741.8	740.0	739.2
P	173.4	183.0	178.4

Table 7.13: Comparison of goodness-of-fit statistics for FM and ZIP models (from Table 7.12, and Poisson model (Model 7, Table 6.6, page 131), Period 2 (m corresponds to the number of parameters in the model).

those for the Poisson model, Model 7, presented in Table 6.6 (page 131). The Poisson model performs similarly, with respect to the goodness-of-fit statistics, to the ZIP model, with the FM model best. The Poisson model contained two climatically based covariates, namely $WindS^{m}_{mean(std)}$ and $South^{m}_{-diff}$. This Poisson model adequately described the relationship between climate and the incidence of SIDS at a monthly level in Period 2, although the model overestimated predicted SIDS for months where one SIDS occurred. The FM model also overestimated the class corresponding to one SIDS per month, as well as the zero class. The predicted number of months where two SIDS occurred was underestimated by the FM model.

With respect to ease of implementation and interpretation of the regression coefficients, the Poisson regression model is preferred over the FM model, for describing the relationship between SIDS and climate in Period 2. Although the FM model is naturally understandable with the concept of underlying subpopulations, in this case the additional complication of model interpretation and computation do not warrant the gains made: component membership in the FM model was defined purely on a seasonal basis, adding little further information to the climate risk profile given by the Poisson model.

Both the ZIP and Poisson models performed similarly with respect to goodness-of-fit statistics (Table 7.13), but the ZIP model adds an extra dimensionality to the SIDS - climate risk profile: increasing humidity increases the probability of occurrence of the event (a SIDS death) in any one month. This increasing humidity relates to a decrease of belonging to the 'perfect state' (no deaths occurring). This additional information highlights the value of mixture models over the more standard Poisson regressions, and in this situation, the ZIP model is preferred over the Poisson model.

7.6 Comparison Between Poisson Mixture Models

A comparison between the five Poisson fixture models, in terms of model fit and profiles of SIDS risk is made utilising the model presented for Period 2. In Period 1, the only model capable of modelling the underdispersion present in the observed SIDS distribution was the EPP model. Therefore, comparing the five mixture model results for Period 1 would be inherently biased. The observed distribution of SIDS counts in Period 2 exhibits some overdispersion (with respect to the Poisson distribution), which all five methods are able to model successfully within their underlying framework.

Model Fit and Profiles of Risk

Table 7.14 details the estimated mean values and proportions of months with no SIDS occurring and months with at least one SIDS death for the five Poisson mixture models presented in Table 7.10 and Table 7.12, alongside the corresponding values observed from the distribution of SIDS counts. These values are presented at both an annual and seasonal level, with the proportions presented for the FM model corresponding to the probability of belonging in component j ($j = 1, 2$). The formulation of the NB model does not involve partitioning of the distribution, therefore no proportion values are reported for this model.

The estimated mean values across all five models are reasonably close to the observed mean, at both an annual level, and across the seasons. The hurdle model overestimates the mean number of SIDS per month in autumn and spring (by 5% in autumn and 6% in spring). In contrast, the NB model underestimates values of the mean in winter (by 3%) and spring (by 1%). Of the five models presented, only the hurdle model correctly estimates the low number of SIDS in summer (observed mean = 1.3 SIDS per month), with the remaining models all overestimating the summer mean. None of the models achieved the necessary inflation of SIDS counts evident in the observed series across winter months (observed mean = 3.7 SIDS per month).

Examination of the proportion of months with no SIDS deaths, and months where at least one SIDS occurred, shows that the hurdle model follows the observed split reasonably successfully, in each case (an-

		Observed	NB	FM	ZIP	Hurdle	EPP
Annual	Estimated mean	2.387	2.369	2.390	2.388	2.390	2.370
	Proportion of zeros	0.147		0.500[a]	0.054	0.146	0.269
	Proportion of non-zeros	0.853		0.500[b]	0.946	0.854	0.731
Summer	Estimated mean	1.275	1.310	1.305	1.311	1.275	1.306
	Proportion of zeros	0.333		0[a]	0.128	0.323	0.269
	Proportion of non-zeros	0.667		1[b]	0.872	0.677	0.731
Autumn	Estimated mean	2.137	2.174	2.183	2.253	2.242	2.207
	Proportion of zeros	0.137		0.926[a]	0.087	0.156	0.069
	Proportion of non-zeros	0.863		0.074[b]	0.913	0.844	0.931
Winter	Estimated mean	3.667	3.549	3.546	3.507	3.511	3.516
	Proportion of zeros	0.020		1[a]	0.0001	0.022	0.014
	Proportion of non-zeros	0.980		0[b]	0.9999	0.978	0.986
Spring	Estimated mean	2.471	2.443	2.524	2.480	2.535	2.452
	Proportion of zeros	0.098		0.074[a]	0.0001	0.081	0.147
	Proportion of non-zeros	0.902		0.926[b]	0.9999	0.919	0.853

Table 7.14: Estimated mean values and proportions of months with no SIDS occurring (zeros), and at least one SIDS occurring (non-zeros), Period 2.

[a] probability of belonging in Component 1.
[b] probability of belonging in Component 2.

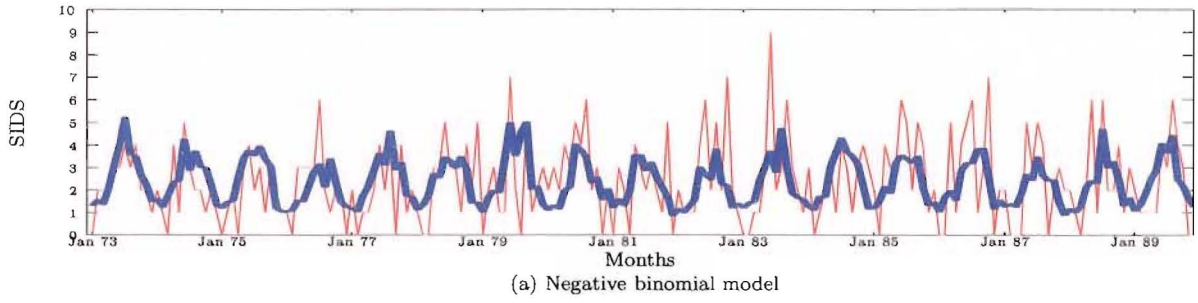
nually, and within each season). The ZIP model consistently underestimates the proportion of months with no SIDS deaths.

Figure 7.19 presents the predicted number of SIDS per month overlaying the observed numbers over time. Each predicted series was calculated by incorporating an autoregressive term (y_{i-1}^m) into the model. The plots are similar over time, with each model following the underlying seasonal pattern inherent in the SIDS distribution. The NB, hurdle and EPP models (Figure 7.19 (a), (d) and (e) respectively) all show a predicted number of SIDS per month which does not fall below one. The FM model appears to follow the peaks and troughs of the SIDS series reasonably successfully.

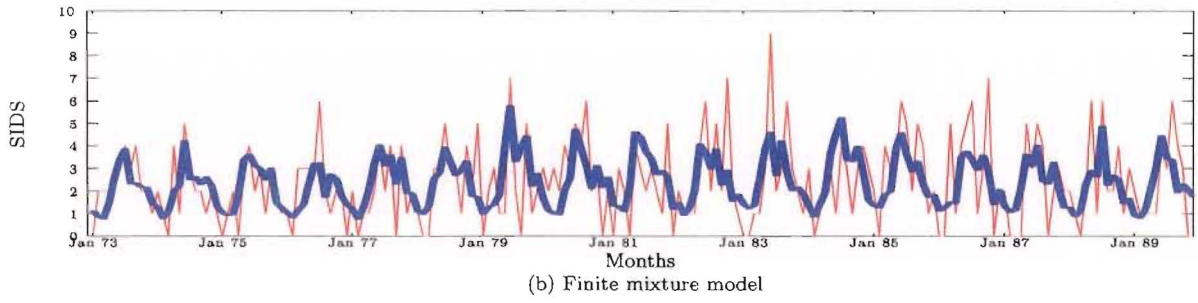
Implementation Issues

When comparing different ways of modelling data, ease of model implementation and identifying parameter estimates, must also be considered, in addition to issues such as model accuracy, appropriateness and predictive performance. Of the five Poisson mixture models presented in this chapter, the negative binomial and zero-inflated Poisson models are available as options in standard statistical software. Inherently, having standard software options for model fitting makes the task of identifying the best covariate structure and calculating parameter estimates easier.

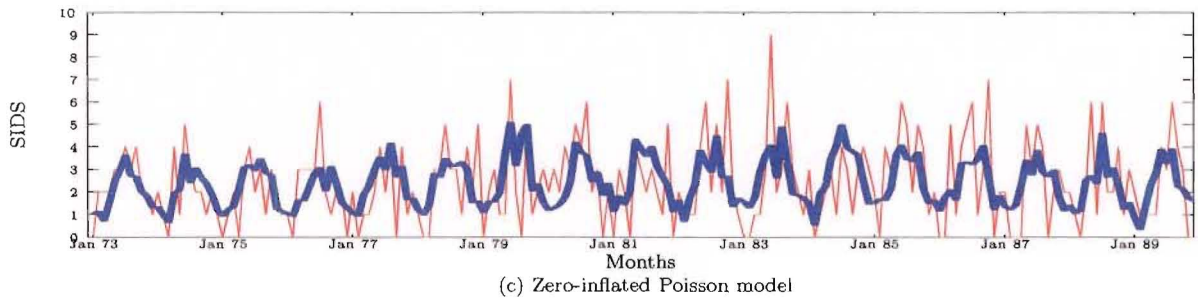
Identifying MLEs in the hurdle model involved a two step process: implementing a logistic regression procedure, available in most standard statistical software packages, and using a specifically constructed Matlab macro to model the truncated Poisson component of the distribution. Similarly, a specifically constructed Matlab macro was necessary to identify parameter estimates in the EPP model. Fortran code was available to calculate the FM model parameter estimates. Like the hurdle modelling, calculating parameter values for the FM model was also an involved process; parameters for the two Poisson components as well as the mixing probabilities needed to be estimated. Time to convergence, and choices of initial values, were potential problems with these less standard modelling techniques (FM, hurdle and EPP). This was not the case when implementing the in-built procedures for the ZIP and NB models.



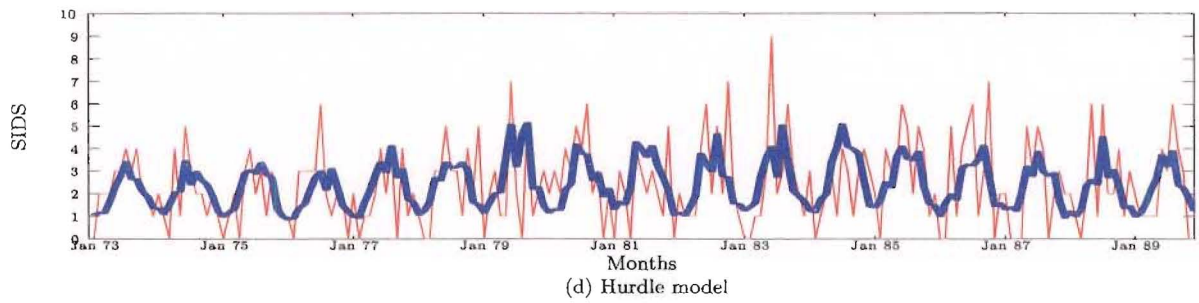
(a) Negative binomial model



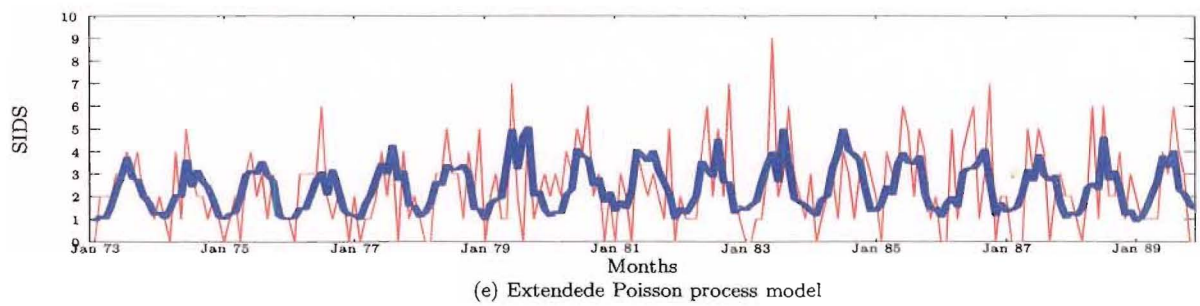
(b) Finite mixture model



(c) Zero-inflated Poisson model



(d) Hurdle model



(e) Extendede Poisson process model

Figure 7.19: Observed —— and predicted ▬ number of SIDS per month for each of the five Poisson mixture models, Period 2.

7.7 Summary and Conclusion

The five methods for modelling count data discussed in this chapter all extend the Poisson distribution, and can successfully incorporate overdispersed data. The EPP model can also model data that is underdispersed with respect to the Poisson distribution. Four of the methods involve a partitioning of the underlying distribution: the FM model into two (different) Poisson distributions, both the ZIP and hurdle models split into zero and non-zero components, while the EPP method models the distinct observation classes.

This section of analysis extends the work of Dalrymple et al. (2003). Dalrymple et al. (2003) examined monthly Canterbury SIDS data from Period 2 (1973—1989) in relation to four monthly climatic variables, alongside variables defining the number of infants at risk of SIDS in any one month, and a seasonality term. The climatic variables involved only relative humidity, temperature, and temperature variants. The analysis undertaken in this thesis initially involved over 2500 daily climatic measures, summarised into 101 monthly climate variables, involving wind speed, direction, and velocity, pressure, rainfall, sunshine, radiation, and dewpoint alongside measures of relative humidity and temperature. This section of analysis based on Poisson mixture models incorporated two further methods alongside those examined by Dalrymple et al. (2003), namely the negative binomial and extended Poisson process models. Results were presented for both Period 1 and Period 2.

Comparing the resulting Period 2 models presented in this chapter with the models presented by Dalrymple et al. (2003) highlights some differences in covariates. The construction of the models presented by Dalrymple et al. (2003) purposely excluded an intercept term, as the NAR variable (describing the number of infants at risk of SIDS) was included specifically as a surrogate intercept term.

The FM model presented by Dalrymple et al. (2003) had one component defined solely by NAR. In contrast, one component of the FM model presented in this chapter (Table 7.10) is defined in terms of seasonality variables. There are necessary differences between the models of Dalrymple et al. (2003) and those presented here with respect to climatic dependencies, as differing climatic covariates were used in each analysis. It is interesting to note that humidity appeared in the Poisson rate component of all three models presented by Dalrymple et al. (2003). Humidity only appeared in the ZIP and hurdle models presented in this chapter, and then it was incorporated in the probability component.

The probability components of the three models presented by Dalrymple et al. (2003) were all structurally defined by one temperature based covariate. In contrast, the probability components of the models presented in this chapter all incorporated seasonality ($\sin(\frac{2\pi t}{12}) + \cos(\frac{2\pi t}{12})$), and in some cases, additional climatic covariates.

The models presented in this chapter are the result of a detailed analysis and a thorough parameter search involving a multiplicity of climate variables. Therefore, they can be considered an improved version of those models presented by Dalrymple et al. (2003).

The EPP model was the only one of the five methods presented in this chapter, to successfully model the slight underdispersion (with respect to the Poisson distribution) seen in the observed distribution of SIDS in Period 1. This EPP model improved on the baseline Period 1 Poisson regression model presented in Chapter 6. The Poisson regression analysis for Period 1 did not find any significant relationships between the incidence of SIDS and climatic covariates, over and above the baseline model form. In contrast, the EPP model included two climatic covariates ($Humid^{dm}_{-diff}$ and $West^{dm}_{+diff}$), enabling it to better capture the risk profile of SIDS.

In Period 2 the resulting models showed similarities. The climatic covariate $WindS^{m}_{mean(std)}$ featured strongly in the Poisson components of all five models, with an increase in $WindS^{m}_{mean(std)}$ corresponding to an increased risk of SIDS. In terms of goodness-of-fit statistics, the FM model was considered best, and highlighted differences in the two underlying Poisson distributions making up the observed SIDS counts.

The ZIP model also successfully related climate to SIDS. Both these models failed to predict the extremes of the observed series.

The resulting Period 2 mixture models are a similar structure to the corresponding Poisson regression model presented in Chapter 6. The baseline model form is similar ($intercept + season + NAR$), and $WindS^{m}_{mean(std)}$ features in both the mixture and Poisson regression models.

Dewpoint based variables featured strongly in the Period 1 models, alongside humidity and wind velocity based variables (see Table 7.4). Humidity is also featured in the structure of some of the Period 2 mixture models (though in a different form), along with wind speed (Table 7.10).

It was interesting to note that in Period 1, the hurdle model performed well with respect to the goodness-of-fit statistics (Table 7.4), yet poorly predicted the probability distribution of SIDS (Figure 7.2). This highlights the necessity of examining model performance through a variety of different measures, and not relying on overall statistics.

The systematic lack of fit at extreme SIDS counts, evident across all five methods, along with the necessary inclusion of an autoregressive term to the models, may indicate some further latent structure involved in the aetiology of SIDS, whether climatically, environmentally, or physiologically based. This needs further investigation.

Chapter 8

In Search of Latent Structure and Non-linearity of Climatic Effects in SIDS

The development of methods specifically designed for regression models involving time series count data has evolved in recent years (see Cameron & Trivedi (1998) or Brockwell & Davis (2002) for an overview). These models essentially fall into two classes: observation-driven models and parametric-driven models.

The conditional distribution of the series of interest, y_t, is defined in terms of lagged observations $(y_{t-1}, \ldots, y_1)$, in addition to the regression covariates, in the observation-driven model. See Berglund & Brannas (2001) or Thyregod et al. (1999), for recent examples of applications of models of this genre.

In contrast, in the parameter-driven model, a latent process is introduced in the conditional distribution of y_t. Zeger (1988) presented a model of this type, where serial correlation in y_t is introduced via a multiplicative latent process, with an underlying autoregressive structure. Both overdispersion and autocorrelation are introduced into the y_t series using this method.

Davis et al. (1999, 2000) have also studied parameter-driven models for time series of observed counts. They present a method for obtaining consistent estimators for the standard errors of the parameter estimates (the asymptotic standard error estimates) in the presence of a latent process, from standard generalised linear modelling analysis. They also examine the issue of testing for the presence of such a latent process.

The dataset examined throughout this thesis involves numbers of deaths in Canterbury resulting from SIDS. This data is a time series of counts, observed over a 32 year period. It is not likely, in the time series scenario, that adjacent observations are uncorrelated. This issue has already been highlighted when modelling the monthly SIDS series, with the necessary inclusion of the y_{t-1} term as a predictor in the regression models presented in both Chapter 6 and Chapter 7. Extending the basic Poisson regression model, to account for serial dependence, may be necessary to obtain legitimate inferences regarding the dependence of SIDS counts on climatic covariates. Section 8.1.1 presents details of Zeger's correlated multiplicative error model, with application to the climate—SIDS scenario examined in this thesis. Section 8.1.3 delineates the asymptotic standard error estimation presented by Davis et al. (1999), and Section 8.1.6 details the test for the existence of a latent process.

The second section of this chapter presents an analysis of the relationship between SIDS and climate using generalised additive models. Generalised additive models are a non-parametric alternative to generalised linear models, and permit non-linear functions of the climatic covariates to be incorporated into the regression model.

Notationally, a subscript 't' will be used throughout derivations and formulations in this chapter, to denote the time series nature of the methods examined. Therefore y_t^m will denote the number of SIDS in month t, and similarly $\mathbf{x}_t^m$ will represent the coefficient vector in month t.

8.1 Time Series Regression using Parameter-driven Models

The assumption, that the serial correlation observed in the data is a result of some unobserved process, is reasonable in the SIDS situation. Campbell (1994) states that "in general a death of one baby does not directly cause that of another", and surmises that the reason for the observed autocorrelation is underlying time ordering inherent in both the dependent and independent data. The two methods presented below, time series regression with multiplicative errors (Section 8.1.1), and asymptotic standard errors (Section 8.1.3), are extensions to the standard Poisson regression model presented in Chapter 6.

8.1.1 Time Series Regression with Correlated Multiplicative Errors

Zeger (1988) introduced a regression model for a time series of counts in which serial correlation was assumed to arise from a latent process. The latent process introduces both overdispersion and autocorrelation into the observed series. Conditional on this process, the observed series is independent. This is a parameter-driven model, which applies quasi-likelihood results to extend the usual generalised linear model framework (McCullagh & Nelder, 1989). This model has been applied by Zeger (1988) to a time series of polio cases, and by Campbell (1994) to relate a daily time series of SIDS counts to short-term temperature trends.

The model formulation below utilises the notation of Zeger (1988), and follows the derivations given by Zeger (1988), and Campbell (1994), along with Cameron & Trivedi (1998).

Parameter Driven Model

Let y_t^m denote the number of SIDS in month t, $t = 1, \ldots, n$ ($n = 59$ for Period 1 and $n = 204$ for Period 2); $\mathbf{Y}_t^m$ is then the vector of monthly responses. Let $\mathbf{X}_t^m$ denote the matrix of monthly covariates, with $\mathbf{x}_t^m = [1, x_{1t}^m, \ldots, x_{kt}^m]'$, where there are a total of k covariates. This is the same situation as described in Chapter 7 with the Poisson mixture models, and in Chapter 6 with Poisson regression. Furthermore, let ϵ_t represent an unobserved stationary process with

$$\begin{aligned} E[\epsilon_t] &= 1 \\ Var[\epsilon_t] &= \sigma^2 \\ cov[\epsilon_t, \epsilon_{t+\tau}] &= \sigma^2 \rho_\epsilon(\tau), \end{aligned} \tag{8.1}$$

where $\rho_\epsilon(\tau) = corr[\epsilon_t, \epsilon_{t+\tau}]$ ($\tau = 1, 2, \ldots$), is the autocorrelation function (ACF) of ϵ_t.

Given both ϵ_t and $\mathbf{X}_t^m$, the (conditional) probability density function of $\mathbf{Y}_t^m$ is independent over t, and of the form

$$P(\mathbf{Y}_t^m | \mathbf{X}_t^m, \epsilon_t) = \text{Poisson}(\lambda_t \epsilon_t), \tag{8.2}$$

where $\lambda_t = \exp(\mathbf{X}_t^m \boldsymbol{\beta}_t)$. The mean and variance (conditional on ϵ_t) are then

$$\begin{aligned} E[y_t | \lambda_t, \epsilon_t] &= \lambda_t \epsilon_t \\ Var[y_t | \lambda_t, \epsilon_t] &= \lambda_t \epsilon_t. \end{aligned} \tag{8.3}$$

The marginal distribution of $\mathbf{Y}_t^m$ (marginal with respect to the latent process ϵ_t only) has conditional mean and variance of the form

$$\begin{aligned}\mu_t &= E[y_t|\lambda_t] \quad = \lambda_t \\ \sigma_t^2 &= Var[y_t|\lambda_t] = \lambda_t + \sigma^2\lambda_t^2.\end{aligned} \tag{8.4}$$

The variance in equation 8.4 above is the same form as the conditional variance of the negative binomial distribution (equation 7.6, page 139), with the degree of overdispersion relative to the Poisson distribution dependent on both λ_t and the corresponding variance of the latent process, ϵ_t.

The autocorrelation function for $\mathbf{Y}_t^m$ is

$$\rho_y(t,\tau) = \frac{\rho_\epsilon(\tau)}{\left[(1+(\sigma^2\lambda_t)^{-1})(1+(\sigma^2\lambda_{t+\tau})^{-1})\right]^{1/2}}. \tag{8.5}$$

8.1.2 Fitting the Parameter Driven Model

A time series analogue of the quasi-likelihood approach (McCullagh & Nelder, 1989) is used to estimate the model parameters, β_t. This incorporates information about the mean, variance and covariance of the observed SIDS series, $\mathbf{Y}_t^m$. Parameter fitting follows the iterative weighted and filtered least squares algorithm presented by Zeger (1988).

The following variables are defined as:

$$V = Var(\mathbf{Y}_t^m) \tag{8.6}$$

$$Z = \frac{d\lambda_t}{d\beta_t}'\hat{\beta}_t + \mathbf{Y}_t^m - \lambda_t \tag{8.7}$$

$$A = \text{diag}(\lambda_1, \ldots, \lambda_n) \tag{8.8}$$

$$D = \text{diag}(\lambda_t + \sigma^2\lambda_t^2) = diag(\sigma_t^2) \tag{8.9}$$

$$C = LD^{-\frac{1}{2}}AX \tag{8.10}$$

where L is an autoregressive filter of the form such that $L\mathbf{Y}_t^m = y_t^m - \alpha_1 y_{t-1}^m - \ldots$. For example if $\mathbf{Y}_t^m$ has an AR(1) structure, then L is defined such that

$$L\mathbf{Y}_t^m = y_t^m - \alpha_1 y_{t-1}^m. \tag{8.11}$$

Let $R(\alpha)$ represent an autocorrelation matrix for a stationary autoregressive process, with parameters α. Then $R^{-1} = L'L$ (Winkelmann, 2000). Also, the following approximations and information matrices are defined as:

$$V \approx V_R = D^{\frac{1}{2}}R(\alpha)D^{\frac{1}{2}} \tag{8.12}$$

$$V_R^{-1} \approx D^{-\frac{1}{2}}L'LD^{-\frac{1}{2}} \tag{8.13}$$

$$I_0 = (AX)'V_R^{-1}(AX)/n \tag{8.14}$$

$$I_1 = (AX)'V_R^{-1}VV_R^{-1}(AX)/n. \tag{8.15}$$

The non-linear weighted least squares estimator for β_t solves the score function (Dobson, 2002) given by

$$\frac{d\lambda_t}{d\beta}'V^{-1}(\mathbf{Y}_t^m - \lambda_t) = 0. \tag{8.16}$$

In the situation where the observations $\mathbf{Y}_t^m$ are independent, V is a diagonal matrix. In the time series situation, V includes off-diagonal terms which depend on the covariance structure of the latent process.

The standard iterative weighted least squares procedure for $\boldsymbol{\beta}_t$ requires the inversion of V. Zeger (1988) showed that this inversion could be avoided by using generalised estimating equations, where V is replaced by an invertible approximation (see equations 8.12 and 8.13). Values of $\boldsymbol{\beta}_t$ remain consistently estimated. (See Zeger (1988) for full details.)

In the iterative procedure used to calculate $\hat{\boldsymbol{\beta}}_t$, the parameter estimates of $\boldsymbol{\beta}_t$, the $(j+1)$th update is given by

$$\hat{\boldsymbol{\beta}}_t^{(j+1)} = (C'C)^{-1}C'(LD^{-\frac{1}{2}}Z), \tag{8.17}$$

where the current parameter estimates $\hat{\boldsymbol{\beta}}_t^{(j)}$ are used to calculate the update.

Standard errors for $\boldsymbol{\beta}_t$ are calculated using the fact that $\sqrt{n}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t)$ is asymptotically normal, with mean 0 and covariance

$$V = I_0^{-1} I_1 I_0^{-1}. \tag{8.18}$$

The matrices I_0 and I_1 are the information matrices presented previously in equations 8.14 and 8.15.

The parameters, σ^2 and $\rho_\epsilon(\tau)$, of the latent process ϵ_t, can be estimated by a method of moments, as shown by Zeger (1988). The resulting equations are

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^n \left[(y_t^m - \hat{\lambda}_t)^2 - \hat{\lambda}_t\right]}{\sum_{t=1}^n \hat{\lambda}_t^2} \tag{8.19}$$

and

$$\hat{\rho}_\epsilon(\tau) = \frac{\sum_{t=\tau+1}^n \left[(y_t^m - \hat{\lambda}_t)(y_{t-\tau}^m - \hat{\lambda}_{t-\tau})\right]}{\hat{\sigma}^2 \sum_{t=\tau+1}^n \hat{\lambda}_t \hat{\lambda}_{t-\tau}}. \tag{8.20}$$

The form of the variance $\hat{\sigma}^2$ above, is not constrained to be positive. Following the method of Campbell (1994), if the variance becomes negative during the iteration process, it is replaced by $\sum(y_t^m - \hat{\mu}_t)^2/(n-1)$.

8.1.3 Asymptotic Standard Errors

Parameter-driven models, such as that presented by Zeger (1988) (Section 8.1.1), are dependent on the correct specification of the correlation structure in the latent process. Yet, identification of this structure in the latent process is not easily achieved using the observed count process. Davis et al. (2000) show that the autocorrelation structure of the latent process dominates the autocorrelation function of the observed count process, highlighting the difficulty in correctly specifying the correlation structure of the latent process. Significant correlation may exist within the latent process, even when little or no autocorrelation exists in the observed count process. Both Zeger (1988) and Davis et al. (2000) note that using the autocorrelations of the observed process leads to underestimation of the true magnitude of the autocorrelation in the latent process.

Davis et al. (1999, 2000) derived the asymptotic distribution of standard generalised linear model (GLM) estimators for the scenario assumed in this chapter: the existence of an autocorrelated latent process underlying the observed series of SIDS counts.

To test the assumption that an autoregressive latent process exists, consistent estimation procedures are needed for the regression parameters. Both Cameron & Trivedi (1998) and Davis et al. (2000) suggest using GLM procedures to generate the regression coefficients, as these remain consistent in the presence of

autocorrelation. The difficulty then lies in attaining consistent standard errors of the regression coefficients, which are needed for correct inferences to be made.

For the situation where a stationary autocorrelated process exists in the mean of the observed count process, Davis et al. (2000, 1999) prove the consistency and asymptotic normality of the GLM standard error estimates. Their results are analogous to established results for the case of linear regression with autocorrelated errors (Kedem & Fokianos, 2002).

Let $\mathbf{x}_t^m = [1, x_{1t}^m, \ldots, x_{kt}^m]'$ represent a vector of covariates, corresponding to month t, and $\hat{\beta}_{GLM}$ the GLM regression estimates. The results of Davis et al. (1999, 2000) show that the asymptotic covariance matrix of the GLM estimators is :

$$Var[\hat{\beta}_{GLM}] = \hat{\Omega}_{I,n}^{-1} + \hat{\Omega}_{I,n}^{-1}\hat{\Omega}_{II,n}\hat{\Omega}_{I,n}^{-1}, \tag{8.21}$$

where

$$\hat{\Omega}_{I,n} = \sum_{t=1}^{n} \mathbf{x}_t^m \mathbf{x}_t^{m\prime} \exp(\mathbf{x}_t^{m\prime}\hat{\beta}_{GLM})$$

$$\hat{\Omega}_{II,n} = \sum_{t=1}^{n}\sum_{s=1}^{n} \mathbf{x}_t^m \mathbf{x}_s^{m\prime} \exp\left[(\mathbf{x}_t^{m\prime} + \mathbf{x}_s^{m\prime})\hat{\beta}_{GLM}\right] \hat{\gamma}_\epsilon(s-t),$$

and $\hat{\gamma}_\epsilon(\tau) = \hat{\sigma}_\epsilon^2\hat{\rho}_\epsilon(\tau)$ is the autocovariance function of the latent process. Equations 8.19 and 8.20 in the previous section give $\hat{\sigma}_\epsilon^2$ and $\hat{\rho}_\epsilon$. (See Davis et al. (2000) or Davis et al. (1999) for a full derivation, including a theorem and corresponding proof, of this result.)

The asymptotic covariance matrix in equation 8.21 above, is composed of two components: $\hat{\Omega}_{I,n}^{-1}$ is the asymptotic covariance matrix resulting from a standard GLM procedure; while the second term in the equation, $\hat{\Omega}_{I,n}^{-1}\hat{\Omega}_{II,n}\hat{\Omega}_{I,n}^{-1}$, corresponds to the added contribution resulting from the presence of the latent process.

8.1.4 Statistical Modelling Methods for Parameter-driven Models

Both the time series regression with multiplicative errors described in Section 8.1.1, and the asymptotic standard errors detailed in Section 8.1.3, are extensions to the standard Poisson regression models presented in Chapter 6. As such, only the best Poisson regression model from Period 2 (see Table 6.4, page 128) are reanalysed with the parameter-driven time series methods, where the existence of a latent process is assumed. These models do not contain the additional autoregressive term included in the final stage of analysis in Chapter 6. This is because the parameter-driven methods inherently account for this autocorrelation within their respective model formulations, with the assumption of the latent process.

Period 1 is not reanalysed, as neither overdispersion nor autocorrelation has been identified as an issue in the regression models presented for these years.

A specific Matlab macro (*Matlab, Version 6.1.0.1989a, Release 12.1; The MathWorks, Inc.*) was written for use in the calculation of parameter estimates assuming a multiplicative error structure. The parameter estimates from the Poisson regression formulation are taken as initial values in the estimation procedure. A second Matlab macro was written to calculate the asymptotic covariance matrix of the Poisson regression coefficients, and corresponding standard errors, as described in Section 8.1.3. The results from these analyses are presented in the next section, and are compared to Poisson regression estimates for the corresponding models.

8.1.5 Results from Parameter-driven Methods

NOTE: Models analogous to those presented in this chapter, but containing both NAR and an intercept are presented in Appendix G.

Table 8.1 presents the regression estimates for Zeger's method of incorporating a multiplicative error structure into the Poisson model. GLM estimates from standard Poisson regression, and the asymptotic standard errors calculated using the method presented by Davis et al. (1999, 2000) are also given. The three sets of parameters are denoted with subscripts 'Z', 'GLM', and 'A' respectively, with results presented for the best models from Poisson regression analyses (Chapter 6). The parameter estimates and standard errors resulting from time series regression with correlated multiplicative errors (Zeger's method, $\hat{\beta}_Z$ and SE_Z) were calculated assuming an AR(1) structure for the dependence in the latent process.

The asymptotic standard errors have increased marginally, from the GLM standard errors, over all four models. This increase has not affected the significance of the Poisson regression parameter estimates ($\hat{\beta}_{GLM}$).

The resulting estimates from time series regression with correlated multiplicative errors differed from those of the Poisson regression analysis, across all models, though no consistent trend was evident. With the exception of Model 3, the standard errors SE_Z, have all increased, and are generally three to four times greater than the corresponding GLM standard errors (SE_{GLM}). These changes in the parameters have impacted on the significance of the covariates in Models 1, 2 and 4: only the seasonal sinusoid is considered significant in the models calculated using Zeger's method. The regression estimates for Model 3 using Zeger's method are only marginally different from the corresponding GLM estimates, but the $WindD^{md}_{NW}$ covariate is no longer considered significant.

8.1.6 Testing for the Existence of a Latent Process

The parameter-driven time series methods presented by Zeger (1988) and Davis et al. (2000, 1999) returned differing results. Little change was found in the asymptotic standard errors compared to SE_{GLM}, while, in contrast, quite different parameter estimates, and corresponding inferences resulted from Zeger's method. These differences may imply a misspecified autocorrelation structure of the latent process, or even that no such latent process exists in the observed count process. The later premise can be formally tested.

Davis et al. (1999) discusses various statistics for testing for the existence of a latent process. Davis et al. (2000) employ the use of a statistic described by Dean & Lawless (1989). The statistic is

$$OD_B = \frac{1}{\sqrt{2n}} \sum_{t=1}^{n} [\frac{(y_t^m - \hat{\lambda}(\mathbf{x}_t^m))^2 - y_t^m + \hat{h}_{tt}\hat{\lambda}(\mathbf{x}_t^m)}{\hat{\lambda}(\mathbf{x}_t^m)}] \sim \mathrm{N}(0,1) \tag{8.22}$$

where h_{tt} is the tth diagonal element of H, defined in equation 6.14 (page 116). This statistic OD_B was first seen in equation 6.17, page 118, as a test for overdispersion (with respect to the Poisson distribution) in the Poisson regression model. The statistic is defined by the observed count process, and the predicted mean from the regression. Therefore, the asymptotically adjusted standard errors will not effect the value of the statistic, and OD_B will be the same value as reported in Chapter 6. That is, $OD_B = -0.396$, implying that there is insufficient evidence to reject the null hypothesis of no latent process. It should be noted that a pure Poisson point process does in fact generate a time series of independent counts (Cox & Miller, 1977), which implies that there exists no underlying structure.

The relationship between overdispersion and latent processes is not clear when applying the test of Davis et al. (2000) for the existence of a latent process. If a time series is overdispersed, a latent process is inherently assumed. Yet, the converse does not necessarily hold. By using a test statistic developed for the use of testing overdispersion, Davis et al. (2000) imply that a latent process exists only in the presence of an overdispersed series.

Model	Parameter	Zeger's method $\hat{\beta}_Z$	SE_Z	Poisson regression $\hat{\beta}_{GLM}$	SE_{GLM}	Asym. SE_A
Model 1	Intercept	0.889	1.413	1.274	0.527	0.675
	$\sin(\frac{2\pi t}{12})$	−0.469	0.162	−0.342	0.068	0.069
	$\cos(\frac{2\pi t}{12})$	−0.435	0.161	−0.441	0.067	0.070
	NAR	−0.0001	0.0003	−0.0002	0.0001	0.0001
	$WindS^{m}_{mean(std)}$	0.200	0.211	0.164	0.072	0.074
	$South^{m}_{-diff}$	0.184	0.243	0.165	0.078	0.078
Model 2	$\sin(\frac{2\pi t}{12})$	-0.511	0.178	-0.337	0.067	0.067
	$\cos(\frac{2\pi t}{12})$	-0.351	0.177	-0.445	0.067	0.068
	NAR	0.0001	0.0001	0.0001	0.0001	0.00002
	$West^{m}_{mean(mean)}$	0.085	0.191	0.114	0.061	0.061
	$WindD^{md}_{NW}$	0.673	2.395	1.578	0.737	0.731
Model 3	$\sin(\frac{2\pi t}{12})$	-0.407	0.069	-0.336	0.067	0.066
	$\cos(\frac{2\pi t}{12})$	-0.446	0.068	-0.446	0.067	0.069
	NAR	0.0001	0.00004	0.0001	0.0001	0.00002
	$Dew^{m}_{max(max)}$	0.045	0.018	0.036	0.018	0.018
	$WindD^{md}_{NW}$	1.229	0.778	1.683	0.774	0.746
Model 4	$\sin(\frac{2\pi t}{12})$	-0.487	0.196	-0.335	0.067	0.066
	$\cos(\frac{2\pi t}{12})$	-0.435	0.197	-0.444	0.067	0.068
	NAR	0.0001	0.0001	0.0001	0.0001	0.00004
	$West^{m}_{max(max)}$	0.049	0.082	0.048	0.024	0.026
	$WindD^{md}_{NW}$	0.800	2.487	1.636	0.746	0.753

Table 8.1: Regression estimates, with corresponding standard errors for the parameter-driven models.

As previously noted in Chapter 7, the incidence of SIDS in Period 2 exhibit slight, but not significant, overdispersion ($\bar{x} = 2.39$ SIDS per month, $s^2 = 3.14$ SIDS per month). Therefore, according to Davis et al. (2000) the Period 2 data does no contain a latent process. Yet, this data exhibits highly significant autocorrelation (AR(1)).

In the parameter driven model presented by Zeger (1988), the latent process introduces both overdispersion and autocorrelation. It is suggested that a latent process does exist in the Period 2 monthly SIDS profile, primarily as a result of the serial correlation evident in the series, not from overdispersion.

This residual autocorrelation is most likely a result of the temporally ordered covariates inability to fully explain the peaks and troughs in the SIDS series, providing evidence of other (unmeasured) factors influencing the incidence of SIDS.

8.2 Generalised Additive Models

Generalised additive models (GAMs) are a flexible class of models, which incorporate non-parametric regression techniques into the GLM framework. GAMs apply smoothing techniques to describe potential non-linear relationships between the independent and dependent variables. GAMs are an alternative to using polynomial regressors, or searching for the correct functional conversion of both the dependent and independent variables (Rawlings, 1988).

In recent years, GAMs have been utilised in a wide spectrum of applications, including real estate

appraisals (Pace, 1998), identifying risk factors associated with horse falls in steeplechase racing (Pinchbeck et al., 2002), and examining the geographical distributions of sardine eggs and larvae (Stratoudakis et al., 2003). GAMs have become an increasingly popular tool in the epidemiology field. For example: examining the prevalence of the use of hallucinogens (Chilcoat & Schuetz, 1996); investigating the association between infant birth weight and standing at work during pregnancy (Ha et al., 2002); and examining whether mite and cat allergen and bacterial endotoxin levels on mothers' mattresses were associated with cord blood immunoglobulin levels in newborn infants (Heinrich et al., 2002). Another common area to utilise GAMs is in examining the effect of weather on various medical outcomes (Braga et al., 2002; Hajat & Haines, 2002). In a similar vein, GAMS have been the most extensively applied method in time series studies of the relationship between air pollution and morbidity, or mortality data (Erbas & Hyndman, 2001; Donoghue & Thomas, 1999; Stieb et al., 2003; Dominici et al., 2002), and have been applied widely in other time series based analyses (Fryer & Nicholson, 2002; Ramesh & Davison, 2002).

This thesis involves relating climatic time series to a series of mortality outcomes, specifically SIDS deaths. GAMs have not, to date, been utilised in the study of SIDS epidemiology, yet the framework of this study follows similar studies involving the successful employment of GAMs (Kassomenos et al., 2001; Goldberg et al., 2003).

This section presents a brief overview of GAMs methodology and estimation techniques, before applying this modelling method to the analysis of the SIDS—climate relationship. The advantage of GAMs, in this setting, is that it incorporates non-parametric adjustments of potential non-linear effects of seasonality measures, trends in the number of infants in the risk population, and climatic profiles. GAMs present a more flexible approach than the fully parametric models of Poisson regression, and Poisson based mixture models presented in Chapter 6 and Chapter 7.

8.2.1 The Generalised Additive Model

This section gives a brief overview of the GAMs formulation, with respect to its implementation for the SIDS—climate time series data. The formative text on GAMs is Hastie & Tibshirani (1990), which covers all aspects of the theory and application of GAMs.

Let y_t^m denote the number of SIDS in month t, $t = 1, \ldots, n$ ($n = 59$ for Period 1 and $n = 204$ for Period 2); $\mathbf{Y}_t^m$ is then the vector of monthly responses. Let $\mathbf{X}_t^m$ denote the matrix of monthly covariates, with $\mathbf{x}_t^m = [1, x_{1t}^m, \ldots, x_{kt}^m]'$, where there are a total of k covariates. This is the same situation as described in the previous section.

Additive Models

Additive models are an extension of linear regression models, where the regression predictors are replaced by non-linear functions of the predictors. Specifically, the linear regression model can be written as

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{x}_1 + \cdots + \beta_k \mathbf{x}_k, \tag{8.23}$$

where $\mathbf{Y}$ is the dependent variable, $\mathbf{x}_i$ represents the k independent variables, and β_i corresponds to the regression coefficients. The additive model then generalises this linear model by substituting a smooth function $f_i(\cdot)$, for the linear predictor $\beta_i \mathbf{x}_i$, giving

$$\mathbf{Y} = f_0 + f_1(\mathbf{x}_1) + \cdots + f_k(\mathbf{x}_k). \tag{8.24}$$

The smooth functions do not assume any distributional relationship between $\mathbf{Y}$ and $\mathbf{x}_i$, and can therefore be considered non-parametric in nature.

Generalised Additive Models

A similar structure to that of GLM (Chapter 6, page 113) defines the GAM: GAMs consist of three components. The first component is a random component which corresponds to the probability distribution of the response variable. As with GLMs, this probability distribution is required to be a member of the exponential family, and specific to the monthly SIDS series, is the Poisson distribution. The second component is an additive component, which defines the smooth functional forms of the covariates. This component enters the GAM as an additive combination of the smooth functions. The final component of a GAM is the link function, which like GLMs, allows the expected value of the response to depend on the predictor defined by the additive component.

The Poisson regression model presented in Chapter 6, can be written as

$$\boldsymbol{\lambda}(\mathbf{X}^m) = \exp\left(\beta_0 + \beta_1 \mathbf{x}_1^m + \cdots + \beta_k \mathbf{x}_k^m\right). \tag{8.25}$$

The GAM analogue is then simply

$$\boldsymbol{\lambda}(\mathbf{X}^m) = \exp\left(f_0 + f_1(\mathbf{x}_1^m) + \cdots + f_k(\mathbf{x}_k^m)\right), \tag{8.26}$$

which has a log-link function.

GAMs are more complex than GLMs, with the additional functionality inherent in the model specification. Yet GAMs preserve the interpretability so valuable in multiple regression analysis, by modelling the regression surfaces as an additive combination of smooth functions.

Semi-parametric Model

A semi-parametric model contains a combination of linear terms and smooth functions. In the Poisson case, this can be written as

$$\boldsymbol{\lambda}(\mathbf{X}^m) = \exp\left(\beta_0 + \beta_1 \mathbf{x}_1^m + \cdots + \beta_j \mathbf{x}_j^m + f_1(\mathbf{x}_{j+1}^m) + \cdots + f_{k-j}(\mathbf{x}_k^m)\right). \tag{8.27}$$

The first j covariates enter the model as linear predictors of $\boldsymbol{\lambda}(\mathbf{X}^m)$, as in parametric GLM modelling, and the final $k-j$ covariates are smooth functions.

Smoothing Functions

The cubic smoothing spline is the smoothing function utilised in the GAM analysis presented in this chapter. Many other choices for the form of smoothing function exist, including kernel smoothers, running means or medians, or local regression smoothers (see Hastie & Tibshirani (1990) for details).

Smoothing functions are an integral part of GAMs, by nature of the GAM formulation. A smoothing function is a means of summarising the nature of an outcome variable ($\mathbf{Y}^m$) by a function of independent predictors $\mathbf{x}_1^m, \ldots, \mathbf{x}_k^m$. It returns an estimate of the trend in $\mathbf{Y}^m$ that is less variable than the response itself. An important aspect of the smoothing function is that no underlying distributional assumptions are made concerning the relationship between $\mathbf{x}_i^m$ and $\mathbf{Y}^m$; the smoothing function has a non-parametric nature.

Cubic Smoothing Spline

A smoothing spline is essentially a compromise between the accuracy of the fit of the function and the degree of smoothness. Given there are n pairs of data (x_i, y_i), the spline $f_j(\cdot)$ minimises:

$$\sum_{i=1}^{n}(y_i - f_j(x_i))^2 + \lambda \int_a^b (f_j''(t))^2 dt \tag{8.28}$$

where λ is a fixed constant, and $a \leqslant x_1 \leqslant \cdots \leqslant x_n \leqslant b$. The first term in equation 8.28 is a sum-of-squares based term, and describes how closely the spline function fits the data. The second term in the above equation penalises for added curvature in the spline function. It can be verified that equation 8.28 has a single, explicit minimiser, which is a natural cubic spline with knots occurring at the unique x_i values (Reinsch, 1967).

The parameter λ is defined as the smoothing parameter, as it controls the smoothness of the curve: large λ values return a smoother curve, with smaller values of λ giving rise to irregular curves. Cross validation is used to choose the smoothing parameter λ, see Hastie & Tibshirani (1990) for details of this method.

8.2.2 Fitting Generalised Additive Models

An interative approach is used to fit GAMs, involving two algorithms: the backfitting algorithm, which estimates the smooth functions; and the local scoring algorithm , which is similar to the iteratively reweighted least squares procedure used in GLMs.

Backfitting for GAMs

Define the partial residual as

$$R_j = \mathbf{Y}^m - \exp(f_0 - \sum_{i \neq j} f_i(\mathbf{X}_i)). \tag{8.29}$$

Then $E[R_j|\mathbf{X}_j] = f_j(\mathbf{X}_j)$, the jth smoothing function. This expectation is utilised to estimate each smoothing function, $f_j(\cdot)$, in turn, given estimates of the remaining functions, $\hat{f}_i(\cdot)$, $i \neq j$. The ensuing iterative procedure is known as the backfitting algorithm (Hastie & Tibshirani, 1990).

The Backfitting Algorithm

STEP 1: Set initial values

$f_0 = \log(E[\mathbf{Y}])$
$f_1^{(0)} = f_2^{(0)} = \cdots = f_k^{(0)} = 0$
$i = 0,$

STEP 2: $i = i + 1$
For each j, $j = 1, \ldots, k$
$R_j = \mathbf{Y}^m - \exp(f_0 - \sum_{i \neq j} f_i(\mathbf{X}_i^m))$
$f_j^{(i)} = E[R_j|\mathbf{X}_j],$

STEP 3: Repeat STEP 2 until convergence is achieved.

Convergence may be specified in terms of the individual smoothed functions, for example: stop the iterations if $|f_j^{(i+1)}(\cdot) - f_j^{(i)}(\cdot)| < tol$ for all $j = 1, \ldots, k$, and a prespecified tolerance value (tol). Another convergence method is to define an overall statistic, and iterate until this does not change. For example, define the residual sums of squares of the Poisson based GAM as

$$RSS = \text{mean}\big[\mathbf{Y}^m - \exp(f_0 - \sum_{j=1}^{k} f_j(\mathbf{X}_j^m))\big]^2, \tag{8.30}$$

then convergence would be defined as $|RSS^{(i+1)} - RSS^{(i)}| < tol$. It is this second convergence criteria which is utilised in the backfitting algorithm implemented in the GAMs analysis of the SIDS—climate data.

The weighted backfitting algorithm is the same as the unweighted case, but the smoothers are weighted. The weights in this application come from the implementation of the local scoring algorithm.

Local Scoring for GAMs

The difference between GAMs and GLMs is that an additive predictor is substituted for a linear predictor in the GLM. The procedure for estimating the additive terms is similar to that of estimating GLMs using dependent variable regression, which is a type of iteratively reweighted least squares (Härdle, 1990). The GAMs estimation is achieved by substituting the weighted linear regression in the adjusted dependent variable regression by a weighted version of the backfitting algorithm described previously. This gives the local scoring algorithm described below:

The General Local Scoring Algorithm

STEP 1: Set initial values

$f_0 = \log(E[\mathbf{Y}])$

$f_1^{(0)} = f_2^{(0)} = \cdots = f_k^{(0)} = 0$

$i = 0,$

STEP 2: $i = i + 1$

(i) Calculate the adjusted dependent variable (Z), predictor (ν), and mean (μ) using values from the previous iteration

$Z = \nu^{(i-1)} + (\mathbf{Y}^m - \mu^{(i-1)})(\partial\nu/\partial\mu^{(i-1)})$

$\nu^{(i-1)} = f_0 + \sum_{j=1}^{k} f_j^{(i-1)}(\mathbf{X}_j^m)$

$\mu^{(i-1)} = \exp(\nu^{(i-1)})$

Form the weights

$w_j = (\partial\mu^{(i-1)}/\partial\nu^{(i-1)})^2 V_j^{-1}$

where V_j is the variance matrix for Y_j.

(ii) Use the backfitting algorithm with weights w_j to fit an additive model to Z. This returns estimated smoothing functions $\hat{f}_j^{(i)}(\cdot)$.

STEP 3: Repeat STEP 2 until convergence is achieved.

In this GAM application, convergence of the local scoring algorithm is defined in terms of the deviance statistic:

$$G^{(i)} = 2\big[L(\mu^{(i-1)}) - L(\mu^{(i)})\big] \tag{8.31}$$

where $L(\mu^{(i)})$ is the Poisson log-likelihood, calculated using parameters from the ith iteration of the local scoring algorithm. The log-likelihood for the Poisson case has been defined previously in equation 6.8 (page 114), and the deviance statistics in this form was first seen in equation 3.11 (page 65). Convergence is then defined as the mean of $G^{(i)}$ failing to decrease on the $(i+1)$th iteration, in other words $|\text{mean}(G^{(i)}) - \text{mean}(G^{(i+1)})| < tol$ for a prespecified tolerance (tol).

The procedure for estimating GAMs, therefore consists of two algorithms with the backfitting algorithm called within each iteration of the local scoring algorithm (STEP 2 (ii)).

8.2.3 Model and Variable Selection in GAMs

The deviance statistic utilised in both model and variable selection in the GAM is defined as

$$G(\hat{\mu}) = 2\big[L(\mu_{\max}) - L(\hat{\mu})\big] \tag{8.32}$$

where $\hat{\mu}$ is the fitted model, and $\mu_{\max}$ corresponds to the parameter value that maximises the log-likelihood $L(\mu)$ over all μ. The deviance given in equation 8.32 above, is a generalised form of that presented in equation 8.31 which was applied in the convergence of the local scoring algorithm.

Deviance in the Model Selection Process

The deviance statistic G is used to choose between competing GAMs, where the model returning the lower value of G is preferred. This is an analogous process to model selection described in previous chapters utilising various goodness-of-fit statistics.

Deviance and Variable Selection

An approximate F-test is calculated for each candidate spline function in the model. This is achieved by comparing the deviance between the full model and the model without the candidate smooth function (Hastie & Tibshirani, 1990).

8.2.4 Assessing the Fit of the GAM

Partial residual plots are examined for any obvious influential points in the data, as well as any evidence of non-randomness which may imply an incorrect fit in the smoothing function (Hastie & Tibshirani, 1990). The partial residual was defined in equation 8.29 as part of the backfitting algorithm.

The autocorrelation function (ACF) of the model residuals is examined for any evidence of serial correlation. Any deviations from the purely random series are highlighted by extreme values. The model residuals for the GAM are defined as $r_i = \mathbf{Y}_i^m - \mathrm{E}[\mathbf{Y}_i^m]$. If autocorrelation is found in the residual series, an autoregressive lag is added to the model.

8.2.5 Statistical Modelling Methods

A sequential approach is taken in building GAMs, where the best baseline model is found, and then the effect of climatic covariates assessed, over and above this baseline model. The smoothing functions add an extra dimension to this process, in that the significance of the smoothing function also needs to be considered. If the smoothing functions of the baseline covariates (season and number of infants at risk of SIDS (NAR)) are not significant, these covariates are permitted to enter the model linearly. This gives rise to a potentially semi-parametric model.

Models are constructed using the following model forms.

Baseline model

(1) SIDS $\propto$ intercept + spline(season)

(2) SIDS $\propto$ intercept + spline(season) + spline(NAR)

Climate model

(3) SIDS $\propto$ intercept + spline(season) + spline(NAR) + spline(climate variables).

As described in Chapter 5, the monthly variable used to describe seasonality in Period 1 is $Temp^m_{meanMA30}$, the mean of $Temp_{meanMA30}$ over each calendar month. In Period 2, seasonality is described using a sinusoid function, composed of sine and cosine curves (with a period of twelve to represent the annual cycles in the seasonal component of the distribution of SIDS). The number of infants at risk of SIDS in any one month (NAR) is also considered in the baseline models.

A baseline model is formed, selecting the best model form from (1) or (2) above. The splines of season and NAR are only included in the baseline model if they are significantly associated with the incidence of SIDS. The additional contribution of the monthly climate variables (Model (3) above) is initially examined

	Variables	Parameter Estimate (se)	Smoothing Parameter (p-value)	Deviance G
Model 1	Intercept	11.342 (8.629)		49.961
	spline($Temp^m_{meanMA30}$)		0.999 (0.673)	
	spline(NAR)		0.999 (0.469)	
Model 2	Intercept	10.331 (8.172)		55.931
	$Temp^m_{meanMA30}$	-0.058 (0.030)		
	NAR	-0.0016 (0.0014)		
Model 3	Intercept	0.955 (0.342)		57.232
	$Temp^m_{meanMA30}$	-0.059 (0.030)		

Table 8.2: Baseline model results for Period 1, GAMs.

individually, over and above the baseline model. After identifying potentially significant climatic covariates, multivariate GAMs are formed by combining multiple climate variables into Model (3). Finally, the model identified as best describing the relationship between climate and monthly SIDS numbers is assessed for its appropriateness and adequacy. This modelling process is completed separately for Period 1 (1968—1972) and for Period 2 (1973—1989). Where necessary, climate variables are deseasoned using the procedure outlined in Section 3.3.2, page 68 (denoted with a superscript 'd' throughout).

The GAM modelling process was performed using SAS (*SAS (R) Proprietary Software Release (8.02)*) through the GAM procedure.

8.2.6 Results from GAM modelling

Period 1 (1968—1972)

Table 8.2 presents three candidate baseline models for Period 1. The first model (denoted Model 1) contains spline functions for both $Temp^m_{meanMA30}$ and NAR. The approximate F-tests show that the smoothing effects corresponding to both covariates are not significant. Therefore the second model in Table 8.2 (Model 2) incorporates both $Temp^m_{meanMA30}$ and NAR as linear effects. The covariate NAR is not considered significant in the presence of an intercept term and hence, is dropped from the baseline model. This leads to the final model in Table 8.2 (Model 3), which consists of an intercept term and $Temp^m_{meanMA30}$. The $Temp^m_{meanMA30}$ covariate is significantly related to the incidence of SIDS ($p = 0.05$), and is describing the seasonal fluctuations in the SIDS profile. Model 3 is taken as the baseline model in the next stage of the GAM analysis. This model is the same as the baseline model P1(1) in the Poisson regression analysis (Chapter 6, Table 6.1, page 120).

The next stage of the GAM analysis involved incorporating each of the climatic variables into the baseline model, Model 3. The climate variables were integrated into the model through spline smooth functions to give models of the form

$$\text{SIDS} \propto \text{intercept} + Temp^m_{meanMA30} + \text{spline(climate)}.$$

No significant relationships were identified for the climate terms, over and above the baseline model of seasonality. This finding is consistent with the results from Poisson regression analysis on Period 1 data (Section 6.5.1, page 119). It implies that SIDS incidence is not related to climate, at either a linear nor at a non-linear level, after accounting for the seasonal profile in the SIDS deaths. (See Section 6.5.1 for a full

	Variables	Parameter Estimate (se)	Smoothing Parameter (p-value)	Deviance G
Model 1	Intercept	1.612 (0.451)		216.313
	spline($\sin(\frac{2\pi t}{12})$)		0.228 (0.721)	
	spline($\cos(\frac{2\pi t}{12})$)		0.235 (0.681)	
	spline(NAR)		1.000 (0.790)	
Model 2	Intercept	1.682 (0.445)		296.928
	NAR	-0.0002 (0.0001)		
	spline2($\sin(\frac{2\pi t}{12})$, $\cos(\frac{2\pi t}{12})$)[i]		0 (< 0.0001)	
Model 3	Intercept	1.619 (0.455)		229.631
	$\sin(\frac{2\pi t}{12})$	-0.317 (0.066)		
	$\cos(\frac{2\pi t}{12})$	-0.448 (0.067)		
	NAR	-0.00018 (0.0001)		

Table 8.3: Baseline model results for Period 2, GAMs.

[i] Thin-plate smoothing spline, see Wahba & Wendelberger (1980) for details.

assessment of the fit of Model 3, and corresponding interpretation.)

Period 2 (1973—1989)

Three candidate baseline models are presented in Table 8.3. The first model (Model 1) incorporates $\sin(\frac{2\pi t}{12})$, $\cos(\frac{2\pi t}{12})$, and NAR through three spline smoothing functions. With p-values ranging from 0.68 to 0.79 for the approximate F-test, these smoothing effects are not considered significant. This is illustrated in Figure 8.1, which shows the spline function corresponding to NAR. This function appears almost linear over the range of the data.

Model 2 (Table 8.3) contains NAR as a linear term and uses a thin-plate smoothing spline to generate a surface defined by both $\sin(\frac{2\pi t}{12})$ and $\cos(\frac{2\pi t}{12})$ (see Wahba & Wendelberger (1980) for details on this-plate smoothers). This spline is a multi-predictor smoother and introduces a single seasonality function into the GAM. Table 8.3 shows that the smoothing parameter corresponding to the 'spline2' term in Model 2 is zero, indicating possible instability in the model. Therefore Model 2 is not considered as a baseline model for Period 2, due to this instability. Also, Model 2 has a deviance statistic which is much higher than the other two candidate baseline models.

The final model in Table 8.3 is composed of purely linear predictors, and corresponds to Model P2(2) from Poisson regression analysis presented in Chapter 6 (Table ??, page ??). Although NAR returns a p-value of 0.07, deeming it only possibly significant, Model 3 is the taken as the baseline model for further GAM analysis of the Period 2 data. The contribution of the additional information made to the model by NAR in terms of describing the underlying trends in the population at risk was considered important when modelling the incidence of SIDS.

Climate Models

The best GAM model for Period 2 is presented in Table 8.4 (Model 4). This model includes $\sin(\frac{2\pi t}{12})$, $\cos(\frac{2\pi t}{12})$, and NAR as linear predictors and $Pres^{m}_{max(max)}$ as a smoothed function. Figure 8.2 highlights the non-linear relationship between the maximum monthly pressure ($Pres^{m}_{max(max)}$) and the incidence of SIDS, found by the GAM analysis. This figure shows that there seems to be a high risk of SIDS around 1015 HPa,

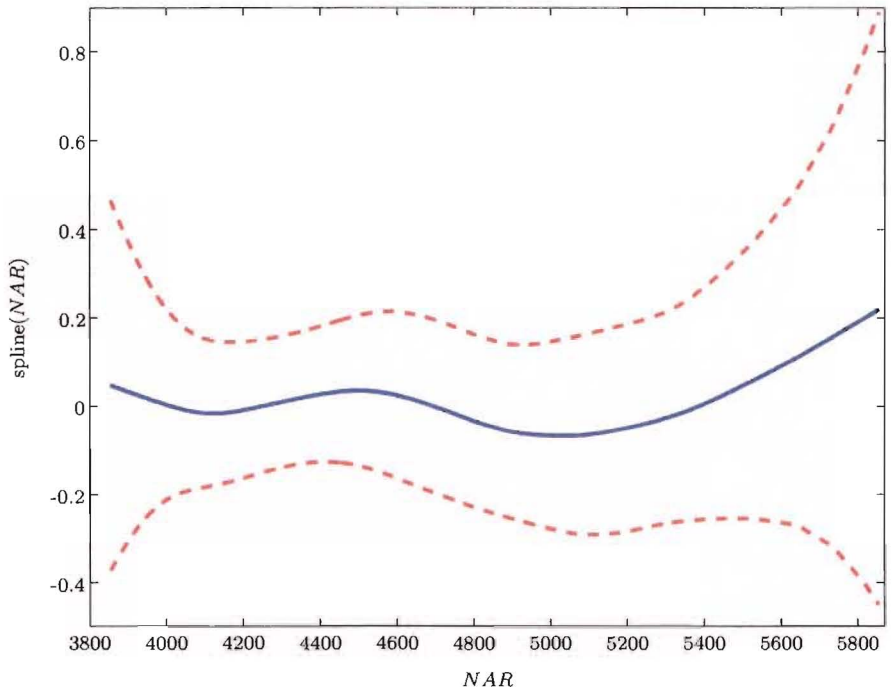


Figure 8.1: Spline function for NAR, Period 2 (- - - represents pointwise 95% confidence intervals).

with a lower risk for lower and higher pressure values.

Residual Diagnostics

Figure 8.3 presents the spline function corresponding to $Pres^{m}_{max(max)}$ in Model 4. In this case the partial residuals are also shown (*). The partial residuals are defined as the fitted values for the spline function plus the model residuals. No obvious trends or patterns are evident in the partial residuals. The point-wise standard error 95% confidence intervals (- - -) are wider at both ends of the range of $Pres^{m}_{max(max)}$, possibly indicating lack of model fit.

Figure 8.4 presents the autocorrelation function (ACF) of the residuals from Model 4. This ACF shows a large value at lag$(t-1)$, implying a possible short-term correlation remaining in the residuals. An autoregressive covariate, comprised of lagged SIDS counts, was therefore incorporated into Model 4.

Details of the GAM containing the autoregressive covariate, given by the number of SIDS per month, lagged by one month (denoted y^{m}_{t-1}), is presented in Table 8.4. This model (denoted Model 5) incorporates y^{m}_{t-1} into Model 4 as a linear predictor. The addition of the autoregressive term significantly reduces the deviance statistic from $G = 217.9$ in Model 4 to $G = 208.1$ in Model 5 ($p = 0.002$) . Figure 8.5 presents the ACF for the residuals from Model 5. This does not show any evidence of a non-random pattern in the time series of residuals. There is a peak at lag$(t-9)$, but this is not considered extreme.

Interpreting the GAM Model

Figure 8.2 highlighted the spline function corresponding to $Pres_{max(max)}$ in Model 4. This gave an indication of the pattern of SIDS risk associated with maximum monthly pressure, over and above seasonality, and the number of infants at risk of SIDS. The lowest SIDS risk occurred at the lowest values of the maximum monthly pressure, and the non-linear relationship defined by the smooth function showed that the greatest risk of SIDS occurs around 1015 HPa.

The predicted SIDS series from the GAM Model 5 (incorporating the autoregressive lagged term) is

	Variables	Parameter Estimate (se)	Smoothing Parameter (p-value)	Deviance G
Model 4	Intercept	9.156 (14.001)		217.873
	$\sin(\frac{2\pi t}{12})$	-0.318 (0.066)		
	$\cos(\frac{2\pi t}{12})$	-0.450 (0.067)		
	NAR	-0.00019 (0.0001)		
	spline$(Pres^{m}_{max(max)})^{i}$		0.999 (0.032)	
Model 5	Intercept	12.219 (14.018)		208.147
	$\sin(\frac{2\pi t}{12})$	-0.420 (0.073)		
	$\cos(\frac{2\pi t}{12})$	-0.509 (0.070)		
	NAR	-0.0002 (0.0001)		
	spline$(Pres^{m}_{max(max)})$		0.999 (0.042)	
	y^{m}_{t-1}	-0.092 (0.029)		

Table 8.4: Climate models for Period 2, GAMs.

[i] $Pres^{m}_{max(max)}$ = monthly maximum of the daily maximum pressure.

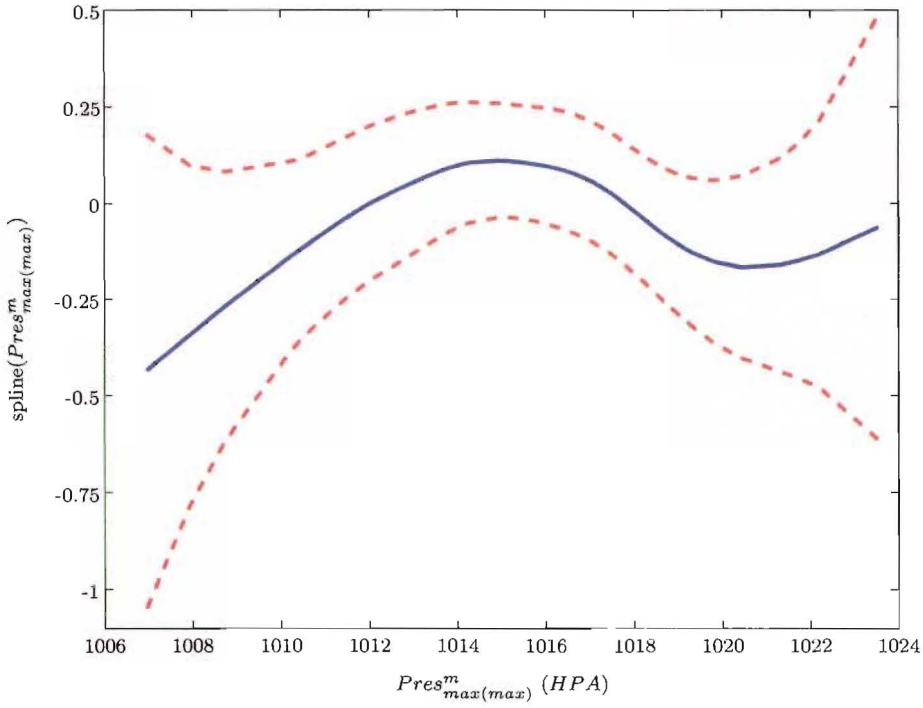


Figure 8.2: Spline function for $Pres^{m}_{max(max)}$, Period 2 (---- represents pointwise 95% confidence intervals).

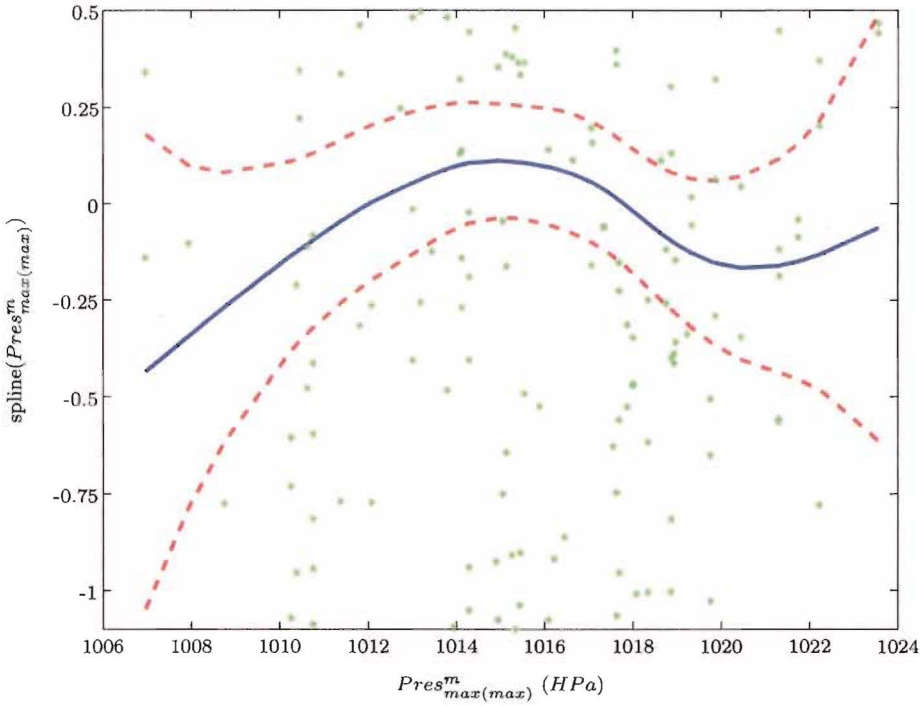


Figure 8.3: Spline function for $Pres^m_{max(max)}$, Period 2 (– – – represents pointwise 95% confidence intervals, * are the partial residuals).

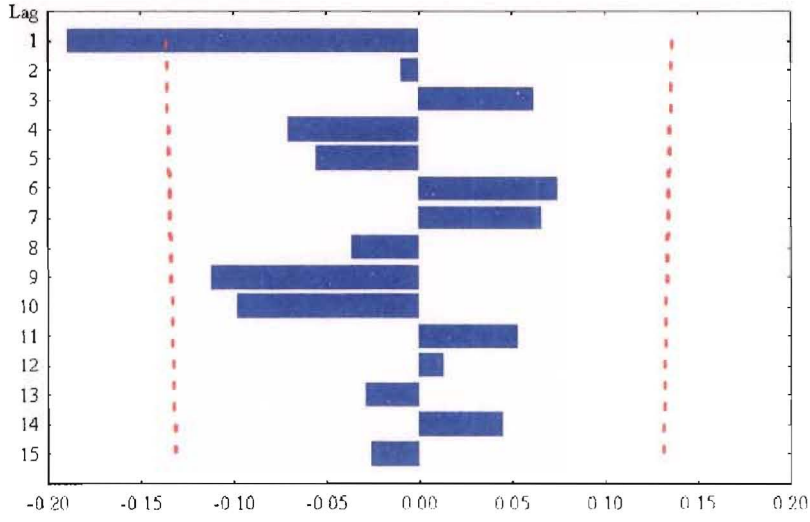


Figure 8.4: Autocorrelation function of the residuals of Model 4, Period 2.

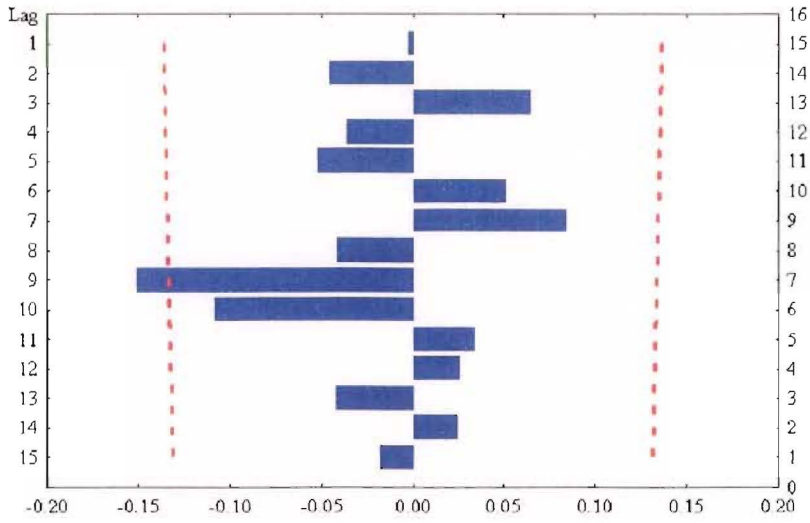


Figure 8.5: Autocorrelation function of the residuals of Model 5, Period 2.

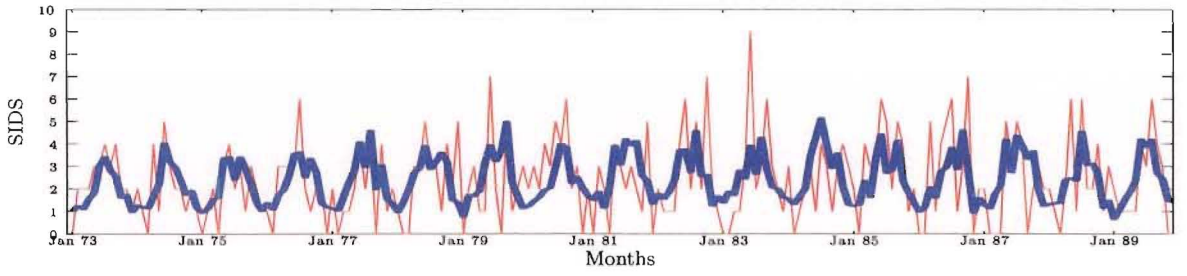


Figure 8.6: Observed —— and predicted ▬ number of SIDS per month for the generalised additive model, Period 2.

presented in Figure 8.6, overlaying the observed series. The GAM appears to follow the seasonal nature of the SIDS series, but the addition of the non-linear function of $Pres^{m}_{max(max)}$ has not aided model prediction. Visually, the GAM model does not closely follow the fluctuations in the observed series, above the seasonality component. The mixture models presented in Chapter 7 (Figure 7.19, page 174) appear to perform better over time than the GAM model.

8.3 Summary and Conclusions

This chapter has presented two distinct formulations, which naturally incorporate the time series profile of SIDS counts. The first section assumed that any serial correlation evident in the SIDS series arose from a latent process. Zeger's (1988) method for estimating the regression parameters in the presence of a latent process was presented, and applied to the SIDS series. This method introduced serial correlation and overdispersion in the observed data via a multiplicative latent process with an underlying autoregressive structure. Davis's (1999, 2000) method for obtaining consistent standard error estimates from standard GLM analysis, when a latent process is present, was also applied to the SIDS—climate model.

	GLM	GAM	Par. Driven
Non-normality	★	★	★
Overdispersion	★	★	★
Non-linearity		★	
Autocorrelation			★

Table 8.5: Summary of modelling aspects accounted for by the three model types.

Results following the development of Davis et al. (1999, 2000) showed little difference between the asymptotic standard errors and the GLM estimates. Zeger's (1988) regression for counts with correlated multiplicative errors resulted in much larger standard errors than the GLM estimates, implying insignificance in some of the model covariates. The assumption of a latent process, which was made in both the correlated multiplicative errors method of Zeger, and Davis's asymptotic standard error calculations, was formally tested. Results showed that no such latent process exists in the observed SIDS series.

The second section in this chapter examined generalised additive models. GAMs have not been previously utilised in the study of SIDS epidemiology. GAMs are a non-parametric alternative to GLMs and permit non-linear functions of the climatic covariates to be incorporated into the regression model.

GAM analysis for Period 1 showed no evidence of non-linear relationships between SIDS and seasonality or the number of infants at risk. No significant climatic functions were related to the incidence of SIDS, over and above the baseline model. GAM analysis for Period 2 also showed the baseline model was best formed by linear predictors of seasonality and NAR. The final Period 2 GAM incorporated the maximum monthly pressure as a spline function. This showed a high risk of SIDS for pressure levels of round 1015 HPa. Campbell et al. (2001) also identified a significant relationship between SIDS and pressure, with a weak positive effect highlighted, where pressure entered the model as a linear predictor. The Period 2 GAM model required the inclusion of an autoregressive covariate, to account for serial correlation in the SIDS series.

Table 8.5 presents a summary of the modelling issues that have arisen from analysing the monthly SIDS time series. The table includes GLMs, GAMs, and parameter-driven models (such as the model presented by Zeger (1988)). A '★' indicates the aspects in the data which the methods incorporate in their formulation. Of the three methods seen in Table 8.5, GAMs are the only formulation capable of incorporating non-linear predictors into the regression model. The parameter-driven model is the only method to directly incorporate autocorrelation, while all three methods can model non-normal data (specifically the Poisson counts in this study), and incorporate overdispersion.

The additional flexibility of the GAM methodology, did not markedly improve the prediction of SIDS counts from analyses presented in previous chapters. The parameter-driven methods were not needed to model the relationship as there was no evidence of a latent process. The regression models presented throughout this thesis, with the inclusion of an autoregressive lag of SIDS counts, are considered adequate to describe the time series inherent in the data set. Although this may seen unusual, a pure Poisson point process does in fact generate a time series of independent counts (Cox & Miller, 1977).

Chapter 9

Discussion and Conclusions

This thesis has investigated the relationship between the incidence of SIDS and climatic patterns using comprehensive statistical analyses. This chapter summarises the information presented throughout the thesis, providing an overview of the major results, a comparison of the statistical methods presented, a synopsis of the climate models defining SIDS risk, a discussion regarding the study strengths and restrictions and finally, an outline of future potential directions in SIDS research.

Relating the methods presented in this thesis to the wider statistical framework

Mixture models have experienced a huge resurgence of interest and appeal over the last two decades. The recent books of Lindsay (1995) Böhning et al. (2000) and McLachlan & Peel (2000), which update earlier books by Everitt & Hand (1981), Titterington et al. (1985) and McLachlan & Basford (1988) attest to this. The enormous advances and frequent application of mixture distributions, (e.g. in meta analysis (Böhning et al., 2000; DerSimonian & Laird, 1986), disease mapping (Schlattmann, 2003; Biggeri et al., 2003; Böhning et al., 2000), genomic research (McLachlan & Peel, 2000; McLachlan et al., 2001, 2003), texture modelling (Grim & Haindl, 2003) and hospital resource utilization (Yau et al., 2003), is due to the fact that mixture models offer natural models for unobserved population heterogeneity (Böhning & Seidel, 2003).

In the time series context, mixture models offer the possibility of approximating nonlinearities in time series, with the advantage that mixtures of simple, perhaps linear components, are often more tractable than more parsimonious nonlinear systems. Mixture models are also important in detecting changes in structure across time and in improving forecasting. The development of mixture based models, for time series data, is mostly concentrated on modelling continuous data as mixtures of normals or as mixtures of ARMA and GARCH models (Bollerslev et al., 1992). The former models are popular in rainfall, solar radiation or seismology studies and the second in the financial literature (Kim & Kon, 1994). Recently Berchtold (2003) proposed a new model for heteroscedastic continuous time series, termed the heteroscedastic mixture transition distribution (HMTD) model. The HMTD model is a mixture of Gaussian distributions and was shown to outperform mixtures of ARMA and GARCH models (Berchtold, 2003). The HMTD, which allows the mean and the standard deviation of each mixture component to be a functional of the past, is a generalisation of the Gaussian mixture transition distribution (GMTD) models of Le et al. (1996) and of the mixture autoregressive (MAR) model of Wong & Li (2000), wherein the mean of each component, only, is written as a function of the past. Like the HMTD, both the MAR and GMTD models are generalisations to the continuous case, of the family of mixture transition distribution (MTD) models first introduced by Raftery (1985) for the modelling of high order Markov chains.

Also, in the mixture of normals context, Glasbey (2001), in a study of solar radiation, proposed a

new form of nonlinear autoregressive (NLAR) time series (Tong, 1990) model, where the joint marginal distributions, at low lag, are assumed to be multivariate Gaussian mixtures. The specific case of univariate Gaussian mixtures, with component membership at time t written as some function of the past, have also been considered in the neural network literature by Jordan & Jacobs (1994) under the general Hierarchical Mixtures-of-Experts (HME) approach. Recently Huerta et al. (2003) looked at the HME application in the context of time series modelling and considered model estimation using a full Bayesian approach. When the mixture component membership is assumed constant over time, the univariate mixture case reduces to the MAR model. The model of Glasbey (2001) is one of a new class of NLAR processes which are time-reversible, with known marginal distributions, and which, in the case where the joint distributions are mixtures of Gaussians, are a subclass of multiprocess dynamic linear models (MDLM) (West & Harrison, 1997), also known as state-space models with switching. Hidden Markov models (MacDonald & Zucchini, 1997) are a special case of MDLM. The NLAR models of Glasbey (2001), unlike most MDLM, have a closed form expression for the likelihood. More recently, work continues in the financial literature on new dynamic mixture models for conditional means and variances, where GARCH and EGARCH (experts) components, can lead to non-normal MDLMs and where the mixture probabilities depend on observed past variables (Dettling & Buhlmann, 2003).

Little work has been done, to date, however which adapts mixture methodology to a discrete time series context. Discrete time series data occur in many applications including health environmetrics, epidemiology, meteorology, econometrics and health policy research. In this thesis five related regression models, based on mixtures of distributions and Poisson processes, have been presented with adaptions to account for discrete count time series data. A preliminary paper gave some suggestions for three of the five models, in the context of a discrete series of SIDS counts and a small subset of climatic predictors (Dalrymple et al., 2003). To date, to the best of the authors' knowledge, there has been only one paper in the published literature that utilises the advantages of mixture models in a discrete count time series context, namely that of Mooney et al. (2003). This recent work fitted mixtures of von Mises distributions (Mardia, 1972) to a particular case study of SIDS rates in the UK between 1983-1998 (data sourced from Douglas et al. (1998)). The monthly SIDS counts, with a range of 70-200 SIDS/month in years prior to 1990 and with a range of 30-65 SIDS/month between 1992 and 1998, were however, assumed continuous. The circular von Mises distribution was deemed appropriate, as the SIDS rates for each year could be considered as being wrapped around a circle. Monthly SIDS rates were analysed separately for each year and covariates were not explicitly incorporated (Mooney et al., 2003). Mooney et al. (2003) report that a topic for future work is that of fitting a mixture of skewed distributions, such as the skew regression models of Batschelet (1981) to this SIDS series.

Only recently, have mixture-based models been extended to accommodate serial correlation in longitudinal studies. For example Booth et al. (2003) extended the negative binomial model to the case of dependent (repeated measures) counts. The dependence was handled by including linear combinations of random effects in the linear predictor; Dobbie & Welsh (2001) have extended the hurdle model to incorporate serial correlation arising from repeated measures count data. They construct general estimating equations for each model component, which incorporate correlation matrices into the component specific maximum likelihood equations; Toscas & Faddy (2003) used transition models to model correlated longitudinal data using the extended Poisson process models of Faddy (1997). There have also been extensions to longitudinal studies exhibiting extra zeros. Hall (2000) adapted zero-inflated Poisson (ZIP) regression models (Lambert, 1992) to an upper bounded count situation by utilising a zero-inflated binomial model and incorporated random effects to accommodate repeated measures data. Dunson & Haseman (1999) extended ZIP regression to a transition model for longitudinal data in an application to carcinogenicity in animal studies.

It is common that data arise also from a mixture of zeros with a continuous distribution, particularly in modelling health claims (Tooze et al., 1994); medical expenditure data (Manning et al., 1987); rainfall

(Katz (1977), Feuerverger (1979), Waymire & Gupta (1981), Stern & Coe (1984), Grunwald et al. (2000), Hyndman & Grunwald (2000) and Grunwald & Jones (2000)); and in modelling audits (Guthrie et al., 1989). In this context, Tooze et al. (2002) proposed a mixed effects, mixed distribution model for repeated measures data with clumping at zero in a study of medical expenditures data. The model assumed the nonzero data arise from a continuous lognormal distribution and included features of the Markov models for time series with mixed distribution of Grunwald & Jones (2000); and of the cross-sectional statistical models of Lachenburch (1976) and Lachenburch (1992). By including correlated random errors, the occurrence and intensity parts of the model were linked and it was assumed unknown whether the zeros arise from the distribution for the occurrence component of the model, or from the intensity component (Tooze et al., 2002). This contrasts the zero-inflated Poisson (ZIP) regression models of Lambert (1992) which contain both these type of zeros. Yau et al. (2002) introduced a zero-augmented gamma random effects mixed model for longitudinal data with many zeros in a study of occupational health claims study. Their decomposition of the likelihood function is, in principle, analogous to the conditional hurdle model approach in modelling discrete counts with extra zeros (see Welsh et al. (1996) and Welsh et al. (2000)). Yau et al. (2002) and Grunwald & Jones (2000) both note that one possible extension of their model is an adaptation containing autoregressive error structure in the random component (Yau & McGilchrist, 1998) or to a transitional model (Grunwald & Jones, 2000). Also for the longitudinal context, Bockenholt (2003) recently introduced a mixture of an integer-valued first-order autoregressive INAR(1) process (Brannas (1994) and McKenzie (1988)) with negative binomial (NB) marginal distributions (Bockenholt (1999) and Joe (1997)) to model repeated measures data on emotional experiences. This INAR-NB factor model with person-specific random effects subsumes various multivariate count models as special cases (Blundell et al. (1995), Cameron & Trivedi (1998) , van Duijin & Bockenholt (1995) and Winkelmann (2000)). In Bockenholt (2003) the order of the INAR process was tested by representing the INAR as a mixed transition distribution (MTD) (Raftery & Tavare (1994) and Le et al. (1996)).

9.1 Summary of Main Chapter Specific Findings

A unique chronological profile of SIDS deaths in Canterbury, over the years 1968—1999, was linked to a comprehensive climatic data set derived from nearly 2.5 million weather observations. The major statistical analyses in this thesis were presented in Chapters 2, 7, and 8. Chapter 1.8 presented an overview of the literature relating SIDS to various climatic patterns. This chapter also described and detailed the comprehensive climatic dataset which was examined in relation to the incidence of SIDS.

Chapter 2: Identifying Change Points in the SIDS series

The chronological profile of Canterbury SIDS deaths was examined in Chapter 2. Three distinct methods were utilised to identify underlying structural shifts in the observed series, namely Survival, Block Bootstrap and Mixture. Previous analyses of SIDS time series have not statistically identified change points in the series, yet to successfully accomplish the aim of this study, the features of the temporal SIDS series needed to be accounted for in the climate modelling process.

The Survival analytic method was a time to event analysis, and therefore accounted for the conditional time aspect of the series. Change points were specified as points of significant change in terms of hazards or SIDS cycles. The Block Bootstrap method used differing definitions of the discriminating statistic to identify change points. This method resulted in both mean SIDS per month, and estimated sample autocorrelation, before and after change points. The Mixture method of identifying change points assumed an underlying heterogeneity in the SIDS series. The distinct groupings were translated into a partitioned temporal map via

a posterior Bayes classification. Of the three change point methods utilised, the Block Bootstrap method was the only method to accommodate the complete monthly SIDS time series, in that block sampling replicated the annual fluctuations in the series. Both the Survival and Mixture methods became insensitive to locating significant change points in the presence of within year seasonal fluctuations.

These three distinct methods of examining the change point problem returned similar results with change points located at 1972 and 1989. This effectively partitioned the SIDS series into three periods:

1968—1972	Period 1
1973—1989	Period 2
1990—1999	Period 3.

All statistical analyses in this thesis therefore incorporated this period effect, which was achieved by viewing the resulting periods as distinct temporal series, and modelling them separately.

Period 3 corresponds to a time when the SIDS numbers had dropped dramatically in Canterbury (from approximately 29 SIDS per year in Period 2 to 9 SIDS per year in Period 3), as a result of 'back-to-sleep' publicity campaigns (detailed in Chapter 1). The temporal series of SIDS counts in this period was shown to be random. Therefore, no further analysis was performed over the years 1990—1999, with respect to relating climate to the incidence of SIDS. Statistically significant seasonal trends, or other climatic patterns could not be successfully identified when SIDS deaths occurred at random time intervals.

Chapter 3: Logistic Regression Analysis of Daily SIDS counts

The SIDS profile, at a daily level was related to the climatic variables using a logistic regression analysis in Chapter 3. The daily observed series was viewed as a dichotomous variable, with the binary outcomes defined as days where at least one SIDS death occurred, and days with no SIDS deaths.

The initial step in the analysis procedure involved identifying the best measure of the seasonal fluctuations in the SIDS series. Seasonality is the only consistently found relationship, when examining SIDS outcomes with respect to various meteorological measures. Yet, there is no standard methodology, or variable, to describe the annual cyclic fluctuations. Various candidate seasonality variables were examined, including indicator functions, sinusoids, and climate based moving averages.

The best measure of seasonality in Period 1 was found to be a thirty day retrospective average of the daily mean temperature ($Temp_{meanMA30}$). In Period 2, the variable that best described the annual cyclic fluctuations in the SIDS series was a first order sinusoid, calculated using an additive combination of both sine and cosine functions, with a 365 day period. The two periods found different measures to best describe the seasonality inherent in the observed SIDS counts, which highlighted the difference in profiles between the periods, identified by the change point findings from Chapter 2.

The second stage of this section of analysis examined each climatic variable individually, for any significant effects on the incidence of SIDS (over and above seasonality). As a data reduction step, this successfully decreased the number of climate variables from approximately 2500 to 125 distinct variables potentially associated with SIDS in Period 1, and a reduction to 171 distinct variables in Period 2.

Significant relationships identified by this univariate-type analysis were regarded with caution, as the weather on any day is defined by complex interrelationships and interdependencies between the individual climate measures. Principal component regression (PCR) is a technique where the correlated variables are replaced by uncorrelated principal components (PCs) in the regression analysis. The principal components are, by definition, not correlated so multicollinearities do not exist between them, and the issue of dependencies within the climatic profile is overcome.

Chapter 4: Principal Component Regression Analysis of Daily SIDS Counts

Climatic components have not previously been related to the incidence of SIDS, and are uncommon in the general epidemiology framework. A general description of the climatic components is given in Table 4.2 (page 86) for Period 1 and Table 4.9 (page 96) for Period 2, with structural differences between the components for each period evident. For example, the first principal component in Period 1, comprising approximately 60% of the variance explained by the PCs, consisted of wind speed and dewpoint. Whereas in Period 2, the first principal component was an amalgamation of temperature, southern wind velocity, rain, sunshine and radiation, and only accounted for 31% of the explained variance.

Principal component regression analysis of the SIDS—climate relationship returned models for Period 1 and Period 2. These models were presented diagrammatically in Figures 4.8 (page 94) and 4.15 (page 103), for Period 1 and Period 2 respectively. These diagrams highlight the hierarchical structure underlying the component built model, and hence the difficulty making practical and meaningful interpretations when such detailed climatic variables are present. The profile of SIDS risk was adequately captured through the PC based covariates incorporated in the PCR models. Yet, the additional dimension involved in the underlying climatic structure of the principal components made interpreting these models a complex issue — creating a climatic summary of what defines a 'high-risk' SIDS day, or on the other hand a 'low-risk' SIDS day, was not possible.

Chapter 6: Poisson Regression Analysis of Monthly SIDS Counts

The next stage of the analysis of the SIDS—climate relationship involved modelling the incidence of SIDS at a monthly level. Information loss was necessarily involved, but was balanced against creating more interpretable monthly climatic covariates.

Poisson regression analysis was used to systematically examine the relationship between climate and the incidence of SIDS at a monthly level in Chapter 6. Poisson regression is suitable for modelling data where the dependent variable consists of count data. This was the case with the number of SIDS per month. The model building process involved incorporating a variable describing the number of infants at risk of SIDS in any month, alongside the seasonality function, to form a baseline model. The effect of monthly climatic covariates on the incidence of SIDS was then examined.

Results from the Poisson regression analysis performed on the Period 1 dataset did not identify any climatic regressors that were significantly related to the incidence of SIDS, over and above the baseline model. The baseline model did not perform well, with regard to predictive ability and residual diagnostics. Underdispersion in the Period 1 SIDS distribution may have been a reason for this.

Two baseline model forms were utilised in Period 2. Both forms contained seasonality and NAR (the number of infants in the risk population for SIDS) with only one baseline model including an intercept term. The omission of the intercept allowed further climatic patterns to be identified. Over these two baseline model forms, sixteen out of the 101 climate variables examined, were significantly related to SIDS incidence (over and above the corresponding baseline model). Most of the sixteen climate variables corresponded to a positive effect on the estimated mean number of SIDS per month.

Chapter 7: Poisson Mixture Modelling of Monthly SIDS Counts

Chapter 7 examined the relationship between SIDS and climate at a monthly level using five different Poisson based mixture model formulations. The Poisson mixture models were to highlight differential climatic effects between months where SIDS occurred and months where there were no SIDS deaths, while retaining the underlying Poisson formulation. Utilising models of this type better captured the climatic dependency of the monthly incidence of SIDS, enabling an enhanced understanding of the relationship between climate

and SIDS deaths. Mixture modelling in this form is uncommon in the general time series framework. The five model formulations were negative binomial, finite mixture, zero-inflated Poisson, hurdle, and extended Poisson processes. They are all extensions of the Poisson distribution formulation, yet characterise the underlying distribution of SIDS counts in different ways.

A description of the underlying distributions in the five mixture formulations in presented in Table 7.1 (page 138). In addition, this table also presents an overview of particular situations where the given models reduce to the Poisson regression model, or one of the other mixture models.

Results from the mixture modelling highlighted relationships between the incidence of SIDS and dewpoint, humidity, wind velocity, and wind speed variables. The analysis presented in this chapter is a major extension of that given by Dalrymple et al. (2003), which was the first study to apply mixture based models to a temporal sequence of SIDS counts. Dalrymple et al. (2003) examined only three mixture models in relation to Canterbury monthly SIDS counts, for the restricted time frame of Period 2 (1973—1989), and with a limited climatic focus (humidity and temperature based climate variables). Extended Poisson Process models have not previously been applied to analyse SIDS risk factors. The analysis presented in this chapter extends the work of Dalrymple et al. (2003) in three ways:

1. Two further Poisson based mixture models are presented (negative binomial and extended Poisson process models);

2. The analysis examines the underlying climatic relationships evident in SIDS counts in both Period 1 and Period 2;

3. A complete climatic profile, involving the 101 climatic covariates defined in Chapter 5, is examined for significant influences on the incidence of SIDS.

The extended Poisson process (EPP) model was the only one of the five mixture methods presented in Chapter 7, to successfully model the slight underdispersion (with respect to the Poisson distribution) evident in the observed distribution of SIDS in Period 1. This EPP model significantly improved on the baseline Period 1 Poisson regression model presented in Chapter 6. The EPP model included two climatic covariates ($Humid^{dm}_{-diff}$ and $West^{dm}_{+diff}$), enabling it to better capture the risk profile of SIDS, compared with the Poisson regression baseline model.

The Period 2 mixture models were similar. The climatic covariate $WindS^{m}_{mean(std)}$ featured strongly in the Poisson components of all five models, with an increase in the variability in wind speed corresponding to an increased risk of SIDS. The finite mixture model was considered best, in terms of goodness-of-fit statistics, and identified differences in the two underlying Poisson distributions defining the observed SIDS counts. The zero-inflated Poisson model also successfully related climate to SIDS.

Chapter 8: Latent Structure and Non-linearity in the SIDS series

Two distinct formulations which naturally incorporate the time series profile of SIDS counts were presented in Chapter 8. The first method assumed a latent process gave rise to potential serial correlation evident in the SIDS series. This latent process introduced both overdispersion and autocorrelation into the observed count process, and conditional on the latent process, the observed series was independent. Two methods were used to estimate regression parameters in the presence of the latent process: assuming a multiplicative latent process with an underlying autoregressive structure, and asymptotic standard error estimates.

The asymptotic standard errors were only marginally different to the GLM estimates. The autoregressive latent process method returned much larger standard errors than the GLM estimates, implying insignificance in some of the model covariates. The assumption of a latent process, which was made in both the correlated multiplicative errors method, and in calculating asymptotic standard error estimates, was formally tested. Results showed that no such process exists in the observed SIDS series.

The second section in Chapter 8 analysed the relationship between monthly SIDS incidence and climate using generalised additive models (GAMs). GAMs are a popular tool for regression type studies of climatic time series and health outcomes, yet to date, have not been utilised in the study of SIDS epidemiology. GAMs are a nonparametric alternative to generalised linear models that allow nonlinear functions of the climatic covariates enter into the regression model.

GAM analysis for Period 1 showed no evidence of non-linear relationships between SIDS and seasonality or the number of infants at risk. No significant climatic functions were related to the incidence of SIDS, over and above the baseline model. GAM analysis for Period 2 also showed the baseline model was best formed by linear predictors of seasonality and NAR. The final GAM incorporated a spline functional of the maximum monthly pressure. This variable indicated a high risk of SIDS for pressure levels round 1015 HPa. This GAM model required the inclusion of an autoregressive covariate, to account for serial correlation in the SIDS series.

9.2 Comparison Between Statistical Methods Applied to the SIDS—Climate dataset

Eight separate statistical models were applied to the SIDS—climate dataset at a monthly level. These models are summarised in Figure 9.1, which shows the formulation of each model, as well as the relationships between the models.

Poisson regression was used as the initial method, which involved a regression model that consisted of Poisson distributed counts. More complicated models were developed that were based on extending the Poisson model: mixture model extensions, assumptions of autoregressive latent processes, and non-linear functionals of the regression covariates. The Poisson distribution assumption is inherent in the other seven models. The relationships between these methods are shown in Figure 9.1. Under certain conditions some of the methods collapse to the Poisson regression method. All of these seven models can incorporate overdispersion with respect to the Poisson distribution.

Tables 9.1 and 9.2 present a comparative description of the eight models used in Chapters 7, 8 and 9. Details presented in the tables cover the type of data each method is suited to, the availability of computer software to implement the method, the interpretability of the resulting model and any additional information.

Naturally, all methods are suited to Poisson data, and can incorporate time series data, while the five mixture methods, parameter driven methods and GAMs are also suitable for overdispersed data. A variety of the methodology presented in this thesis has been developed for distributions other than the Poisson, including a multivariate normal FM model (Jones & McLachlan, 1991), and a geometric hurdle model (Mullahy, 1986).

The software packages SAS and/or Stata contain standard functions that can estimate Poisson, NB, ZIP and GAM models. Fortran code is available for the FM model, and S-Plus code for the EPP models. Implementation of the parameter driven methods and the hurdle model requires programmes to be written for model estimation.

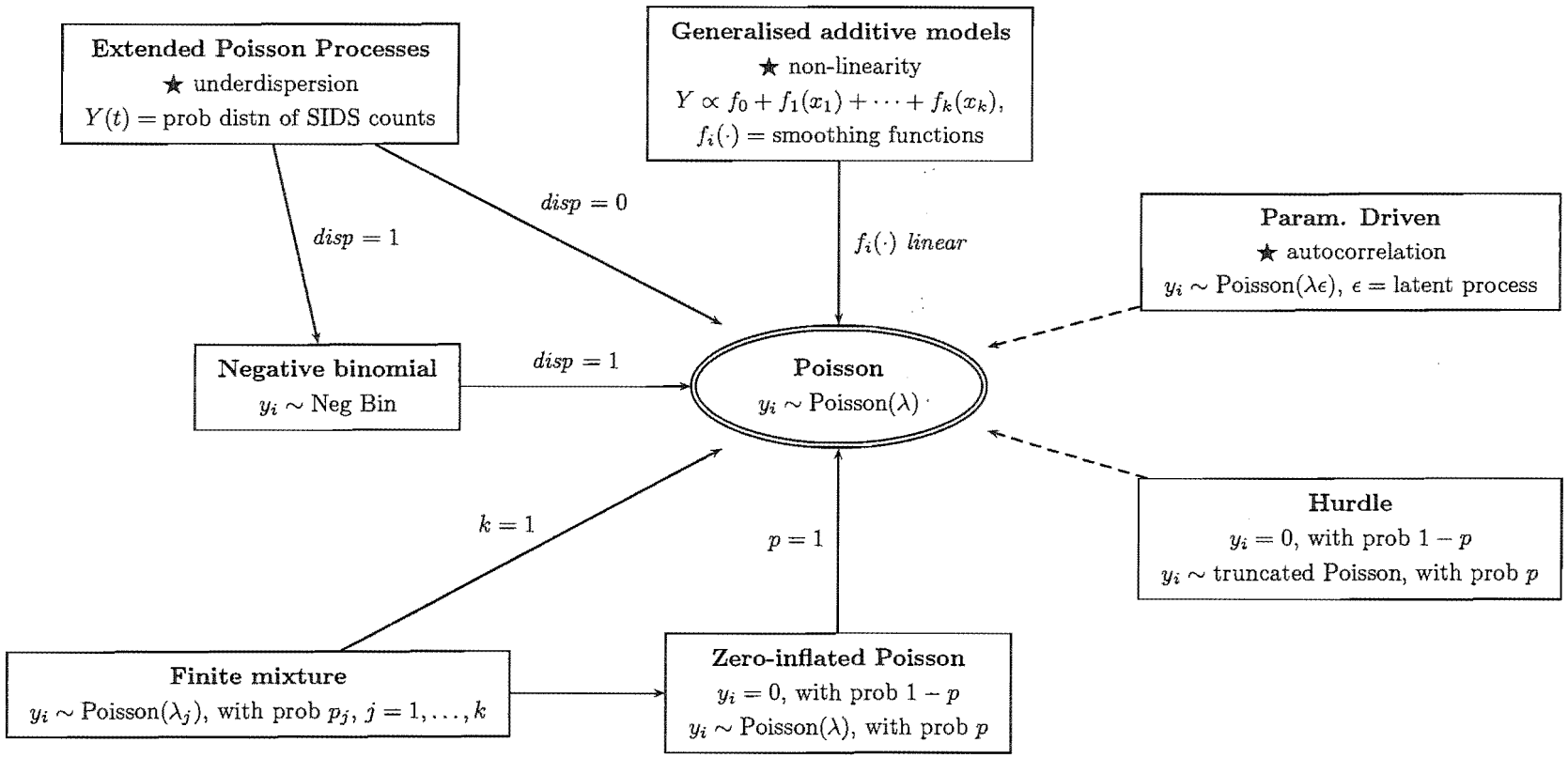


Figure 9.1: Formulation and relationships between statistical methods applied to the monthly SIDS series.

Of the eight models, Poisson, NB and the parameter driven methods generate models that are easily interpreted with respect to quantifying the affect of various covariates on the risk of SIDS. The FM model involves a hierarchical approach to interpreting coefficients: initially the probability of component membership, and secondly, conditional on this membership, coefficients can be interpreted in terms of the Poisson distribution describing that component. Neither the EPP nor GAM models result in a 'nice' interpretation of the regression coefficients.

Some of the eight models have additional features including the FM model, which has the added feature of the subcomponents giving an indication of the latent variable causing the heterogeneity in the data. The hurdle model can indicate both reasons for occurrence of the event of interest (zero/non-zero outcome), as well as examining the risk factors associated with the event.

All of these models are valid in a variety of situations, depending on factors such as the type of data being studied, availability of statistical software, computer programming knowledge, the purpose of the study and what the ideal outcomes (in terms of information gleaned from the models) are. Table 9.3 details a variety of scenarios alongside the method recommended and justification for this choice.

For the Period 2 SIDS data presented in this thesis, the best model (as judged by goodness-of-fit statistics) was the FM model, followed by both the ZIP and Poisson models. With respect to ease of implementation and interpretation of the regression coefficients, the Poisson regression model is preferred. Although the FM model is intuitively understandable with the concept of underlying subpopulations, in this case the additional complexities of model interpretation and computation do not warrant the gains made: component membership in the FM model was defined purely on a seasonal basis, adding little further information to the climate risk profile given by the Poisson model.

With respect to modelling the relationship between SIDS and climate in Period 1, the best method was the EPP model. This method was the only one able to incorporate the slight underdispersion present in the Period 1 data, and the additional climate relationships identified as compared to the Period 1 Poisson model, give an indication of areas for further research. The disadvantages of the EPP model are that quantifying risk in terms of covariates is difficult, and the model cannot be implemented in standard software packages.

Additional Model Choices

A class of Markov zero-inflated Poisson regression models for a time series of counts has been proposed by Wang (2001). These models fall into the category of observation-driven models with the changes in the frequency distribution described by an underlying two-state Markov chain. A second model based on Markov processes is presented by Park & Basawa (2002), who discuss finite mixtures of Markov processes.

Barry & Welsh (2002) have extended the GAMs methodology to the hurdle scenario described in Chapter 7, Section 7.1.4. They present models that provide a greater flexibility than the hurdle model. Specifically, a two-step approach is utilised (as with the hurdle) but each component is modelled as a GAM, where previously a generalised linear model framework was used.

Various other techniques can be used for analysing time series of counts including artificial neural networks (Berry et al., 2002), Bayesian methods (Oh & Lim, 2001; Ibrahim et al., 2000), and transitional regression models (Brumback et al., 2000).

These emerging techniques are outside the scope of the study, but will be considered for future research and extensions.

Method	Data type	Implementation
Poisson	• Poisson count data	• Available in major statistical software packages including SAS and Stata. Also available as part of the 'glmfit' function in Matlab.
NB	• Overdispersed w.r.t. Poisson.	• Available in major statistical software packages including SAS and Stata.
FM	• Method has also been applied to data with distributions that are not Poisson, including Binomial (Aitkin, 1999) and Normal (Jones & McLachlan, 1991). • overdispersed w.r.t. Poisson	• Fortran code available for download from web. Need access to Fortran compilier and basic knowledge of Fortran to implement.
ZIP	• Overdispersed → zero-inflated Poisson count data. • Zero-inflated NB (Succi et al., 2003).	• Available as a function in Stata. • S-Plus code for ZIP models available to download from *www.maths.uq.edu.au/ hmp*
Hurdle	• Method can be applied to combine Bernoulli distn with any truncated count distn, including NB and geometric models (Mullahy, 1986; Welsh et al., 2000).	• Modelled in two parts: PART I: logistic model therefore can use any standard statistical software that contains a logisistic regression procedure (SAS, Stat, S-Plus etc). PART II: truncated (Poisson) model, ML algorithm needs to be written and implemented. As far as author is aware there is no standard software for fitting truncated models.
EPP	• Can model both over and underdispersed count data.	• S-Plus code available for EPP models from *www.maths.uq.edu.au/ hmp*
Param. Driven	• Specific to Poisson time series.	• Algorithm needs to be programmed for implementation; no standard software available as far as the author is aware.
GAMs	• Prob distn required to be a member of the exponential family.	• Major statistical packages starting to implement GAMs software, including SAS (version 9 onwards) and S-Plus.

Table 9.1: Comparison of the data catered for, and the software available for implementation, of the eight statistical methods presented.

Method	Interpretability	Added value
Poisson	• Models are easily interpreted in terms of mean risk profile, with respect to the parameters in the model.	
NB	• Interpretation same as in Poisson regression.	
FM	• Regression coefficients are interpreted w.r.t. mixing probabilities (determing component membership) and the rate functions. Conditional on component membership, interpretation within each component simply reduces to the Poisson regression case.	• Underlying interpretation of sub-components.
ZIP	• Parameters related jointly to the probability of a zero occurrence, or the mean number of SIDS.	
Hurdle	• Parameters can be interpreted separately with prob zero/non-zero and given a non-zero, the mean risk profile.	• Distinction between zeros and non-zero outcomes; can focus on two things when modelling.
EPP	• Covariates do not affect the dispersion parameter, only the gradient coefficient. • Interpretation w.r.t. trends (increasing parameter increases risk) OK but quantifying risk difficult c.f. Poisson.	• Can incorporate both under and over dispersed data.
Param. Driven	• Interpretation same as in Poisson regression.	• Naturally incorporates time series.
GAMs	• The nature of the non-linear functions utilised in GAMS often makes a direct interpretation of coefficient risk difficult, trends can be noted, but not easily quantified.	• Non-linear regressors.

Table 9.2: Comparison of the interpretability, and added value, of the eight statistical methods presented.

Data type and study aim	Method recommended	Why?
• Overdispersed Poisson data, may be a latent variable causing overdispersion.	FM	The partitioning into subcomponents may highlight relationships which bring to light the reason for the component groupings.
• Underdispersed Poisson data.	EPP	Only method capable of modelling underdispersed data.
• Poisson data (time series) – simple regression model required.	Poisson	Easily interpreted and implemented (provided no residual serial correlation).
• Poisson data (time series) – residual serial correlation exists.	Param. Driven	Easily interpreted.
• Poisson data – additional insights into underlying dependencies desired.	FM, Hurdle	These methods have the ability to generate further understanding of complex relationships.
• Zero-inflated Poisson data.	ZIP, NB	Available in standard software; methods specifically designed for zero-inflated data.

Table 9.3: Suggested methods to use for various scenarios.

	Summer			Autumn			Winter			Spring		
$Temp^{m}_{meanMA30}$	16.43			14.31			5.89			10.57		
$Humid^{dm}_{-diff}$	0	−3	−6	0	−3	−6	0	−3	−6	0	−3	−6
E[SIDS/month]	1.09	1.37	1.70	0.81	1.03	1.29	1.84	2.27	2.78	1.30	1.62	2.01
$West^{dm}_{+diff}$	0	2	4	0	2	4	0	2	4	0	2	4
E[SIDS/month]	0.83	0.47	0.26	1.09	0.63	0.35	1.51	0.88	0.50	1.43	0.84	0.47

Table 9.4: SIDS risk for varying levels of climatic predictors, by season, Period 1.

9.3 Climatic Profile of SIDS Risk

The overall aim of this thesis has been to describe SIDS risk with respect to weather patterns. An overall, period specific, profile of what constitutes a high SIDS risk and low SIDS risk month is presented in Figures 9.2 and 9.3 across the methods applied in Chapters 6, 7, and 8. Underpinning SIDS risk is the consistent seasonal pattern of deaths, meaning a high risk of SIDS in colder months, with a corresponding low risk of SIDS in warmer months. This is evident in the models presented.

Period 1 (1968—1972)

Figure 9.2 presents the model specific climatic influences on SIDS risk in Period 1. Of the five mixture models presented Chapter 7, only the extended Poisson process model successfully captured the slight underdispersion (with respect to the Poisson distribution) evident in the Period 1 series. Hence, this is the only mixture model included in Figure 9.2. The parameter-driven methods presented in Chapter 8 were not applied to the Period 1 data, as neither overdispersion nor autocorrelation had been identified as an issue for this time period.

Both the Poisson regression model and the GAM did not identify any significant relationships between

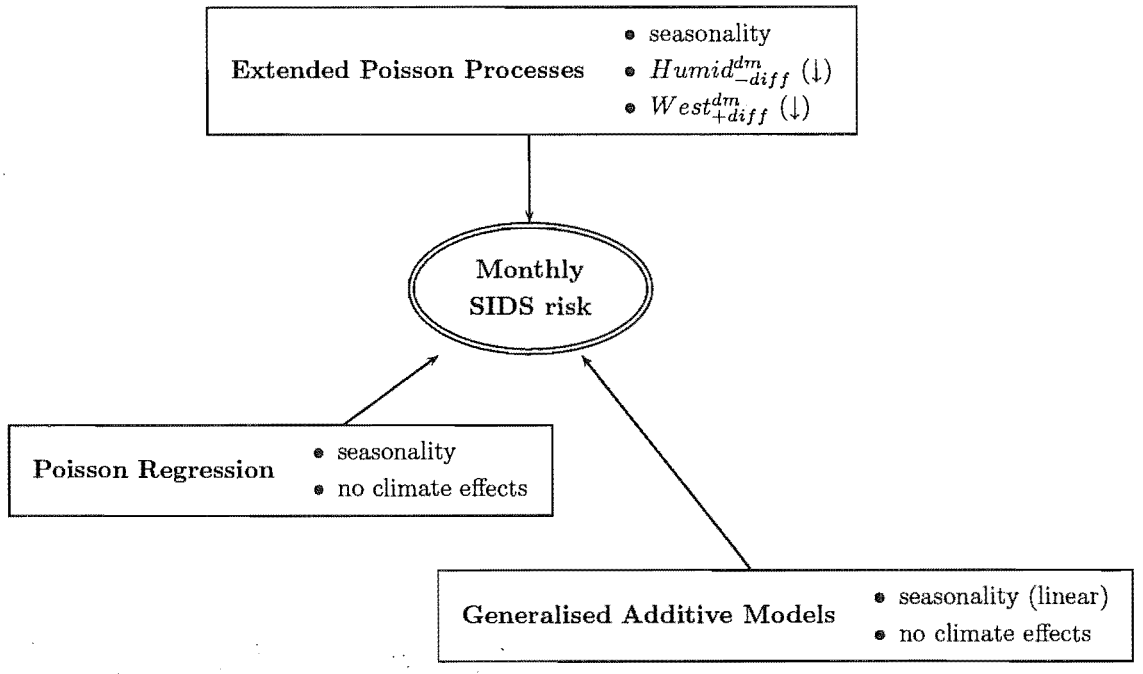


Figure 9.2: Method specific climatic influences (direction of risk) on SIDS risk, Period 1.

SIDS and climate, over and above the seasonal component. The application of the Poisson mixture modelling methods to the Period 1 dataset highlighted some significant climate relationships. The extended Poisson process model presented in Figure 9.2 contains both $Humid^{dm}_{-diff}$ and $West^{dm}_{+diff}$.

The variable $Humid^{dm}_{-diff}$ describes the relative humidity on $day0$ as compared to what it was over the previous seven days, averaged for the month over days where $Humid_{meanday0}$ was less than the respective weekly average. Similarly, the variable $West^{dm}_{+diff}$ describes the strength of the western component of wind velocity on $day0$ as compared to what it was over the previous seven days, averaged for the month over days where $West_{meanday0}$ was more than the retrospective weekly average. In general terms, a month with values of $Humid^{dm}_{-diff}$ or $West^{dm}_{+diff}$ close to zero implies that humidity or western wind velocity has remained reasonably constant, while a month with large values of $Humid^{dm}_{-diff}$ or $West^{dm}_{+diff}$ (in an absolute sense) indicates there have been more extreme shifts away from the average. Increasing values of both $Humid^{dm}_{-diff}$ or $West^{dm}_{+diff}$ correspond to a decreased risk of SIDS, as highlighted in Figure 7.8.

The SIDS risk for Period 1 is quantified in Table 9.4, for both climatic risk factors, by season. The expected number of SIDS per month for the varying levels of $Humid^{dm}_{-diff}$ was calculated using the seasonal average values of $West^{dm}_{+diff}$, and conversely, calculating the SIDS risk for varying levels of $West^{dm}_{+diff}$ was achieved using seasonal averages of $Humid^{dm}_{-diff}$.

Table 9.4 shows that SIDS risk increased by 57% for a decrease from 0% to −6% in $Humid^{dm}_{-diff}$ in summer and by 51% for the same decrease in winter. An increase in $West^{dm}_{+diff}$ from 0 to 4 knots in both summer and winter results in a three-fold increase in the risk of SIDS.

Overall, the climatic models for Period 1 show that the lowest risk of SIDS in Canterbury occurred in summer, in months when both humidity and western wind velocity were reasonably stable. In contrast, the highest risk of SIDS in Canterbury in Period 1 occurred in winter, when the monthly values of both humidity

and western wind velocity recorded extreme shifts away from the average.

Period 2 (1973—1989)

Figure 9.3 presents the method specific climate influences on SIDS risk in Period 2. For each method, the model considered best is detailed. Alongside seasonality, the number of infants at risk of SIDS in any one month (NAR) characterised the baseline models. With the longer time frame of Period 2 compared to Period 1 (204 months compared to 59 months), the changes in the profile of the numbers of infants at risk over this time have a significant influence over the outcome of interest.

The climate effects, over and above the baseline model, are presented in Figure 9.3. This shows $WindS^{m}_{mean(std)}$ (monthly average of the daily standard deviation of wind speed) as a significant climatic covariate in most of the models. With the extended Poisson process, negative binomial and finite mixture models, $WindS^{m}_{mean(std)}$ is the sole climatic variable, while in the Poisson regression, hurdle, and zero-inflated Poisson models $WindS^{m}_{mean(std)}$ is combined with other climatic effects to make up a multivariate risk profile. In the Poisson regression model, $WindS^{m}_{mean(std)}$ is combined with $South^{m}_{diff}$ (monthly average when the southern component of wind velocity was lower on $day0$ than it had been over the past week), while in both the hurdle and zero-inflated Poisson formulations, $Humid^{m}_{min(min)}$ (monthly minimum humidity) is related to the probability of a zero or non-zero outcome. The parameter-driven model returned a different form, with $Dew^{m}_{max(max)}$ (monthly maximum dewpoint) and $WindD^{dm}_{NW}$ (percentage of days in the month where the north-west wind direction predominated) making up the set of significant regressors. The GAM includes $Pres^{m}_{max(max)}$ (monthly maximum pressure) as a non-linear predictor of SIDS. In every case, the linear climate effects are positively related to the incidence of SIDS. The non-linear function of $Pres^{m}_{max(max)}$ in the generalised additive model indicated that the highest SIDS risk occurred with mid-range values of $Pres^{m}_{max(max)}$. Overall, the GAM model did not appear to perform as well as the mixture models, with respect to accurately predicting the number of SIDS per month.

The climate variable $WindS^{m}_{mean(std)}$ is defined as the monthly mean of the daily standard deviation of wind speed. Low values of $WindS^{m}_{mean(std)}$ correspond to months where the daily wind speed was reasonably consistent, whereas, high values of $WindS^{m}_{mean(std)}$ relate to months where the daily wind speed varied widely. In general, using this common covariate of $WindS^{m}_{mean(std)}$ to relate back to the risk of SIDS, a month where SIDS risk was low in Period 2, occurred in summer, when the daily wind speed had been reasonably consistent. In contrast, a month in winter where the daily wind speed varied widely corresponded to a high risk of SIDS.

Impact of these findings on current SIDS research

The advantage of the detailed modelling presented in this thesis has been the ability to highlight potential influences of climate variables other than the standard temperature profile. This provides justification for further detailed research into climate of the post back-to-sleep SIDS profile.

The differing seasonal profile between Period 1 and Period 2 may be related to changes in the Southern Oscillation index (ENSO) (Wolter & Timlin, 1998): Period 2 has been identified as generally warmer and dryer than Period 1. It is hypothesis that the earlier peak in SIDS risk in Period 2 (late autumn/early winter as compared to a peak in late winter/early spring in Period 1) could be indicative of accumulated heat effects, possibly leading to overheating of infants (Wells, 1997)

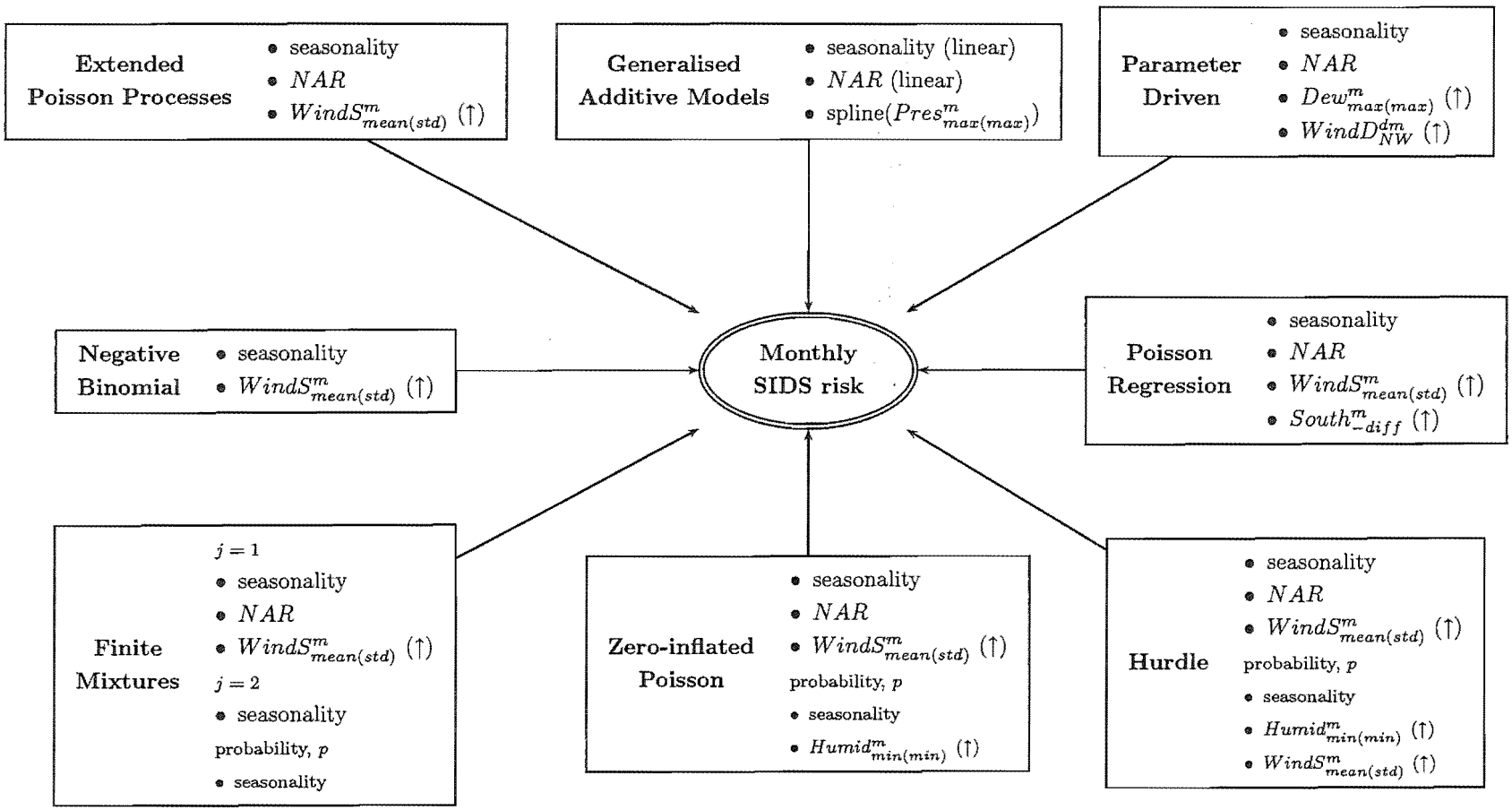


Figure 9.3: Method specific climatic influences (direction of risk) on SIDS risk, Period 2.

The statistical methods in this thesis allude to possible causal pathways relating climate as mediating effects on parental behaviour, which in turn, have an impact of the risk of SIDS. For example, wind has long been associated with ill health. The warm, dry alpine winds, or föhn winds, such as Canterbury's Norwester, have been related to cardiovascular problems, migraines, allergies, and general increased irritability (Verhoef et al., 1995; Miric & Rumboldt, 1993; Piorecky et al., 1997; Carinanos et al., 2000). It is of interest to note the predominance of significant wind variants (whether wind speed or wind velocity) in the modelling results for both Period 1 and Period 2. It is not surprising that variations in wind may have an effect of SIDS; wind variations, which unlike pressure changes, are easily perceived by humans, and may effect parental behavioural patterns or care practices that in turn alter the risk of SIDS. Similar hypotheses in regard to effects of increased temperature (or dewpoint) with sustained overwrapping have also been presented in the literature (Dalrymple et al., 2003; Wells, 1997).

The relationships presented resulted from a detailed exploration of SIDS and a multitude of climate based variables. As such, some significant relationships, though biologically plausible, may in fact be a result of type I errors. Yet, there is extensive debate in the biostatistical and epidemiological literature about when (if ever) adjustment for multiple testing is justified, and also the best way to adjust the p-values, for example Aickin & Gensler (1996); Perneger (1998). For confidence in the effect of the climate relationships identified in this study on the incidence of SIDS, further confirmatory analyses needs to be performed on similar datasets (Bender & Lange, 2001).

The detailed modelling presented in this thesis has necessarily concentrated on the relationship between SIDS and climate prior to the dramatic reduction in SIDS following the back-to-sleep campaigns. The resulting relationships cannot be assumed to hold in the post back-to-sleep environment, where there is already debate on whether the seasonal profile of SIDS incidence still exists (Wigfield et al., 1994; Gilman et al., 1995; Julious, 1997).

9.4 Study Strengths and Restrictions

The strengths of this study can be categorised into two distinct aspects: those specific to the dataset, and those specific to the statistical techniques utilised.

The three main strengths relating to the data are as follows:

1. The SIDS deaths information used in this thesis is the result of a retrospective complete ascertainment study conducted by the Community Pediatric Unit. Autopsy and pathology reports were reviewed for all postneonatal deaths occurring in the Canterbury region, over the period of the study. The SIDS profile is unique in two aspects: firstly, it is unusual to have a series that spans such a long time period (32 years); and secondly, through the nature of the data collection method, the observed SIDS series does not suffer from inaccuracies as a result of changing diagnostic policies and disease category classifications.

2. The meteorological dataset utilised throughout this study consists of hourly weather measurements for the 32 year period of the study. The size and accuracy of this data set enabled an accurate determination and creation of approximately 2500 daily climatic variables based on temperature, wind direction, speed and velocity, wind chill, humidity, pressure, rainfall, sunshine, solar radiation, and dewpoint. Comprehensive day, within day, and between day measures and lags were defined, and related to the incidence of SIDS. Many climatic variables included in this data set had never been examined in relation to the incidence of SIDS, including dewpoint, wind chill and wind velocity.

3. The SIDS deaths were localised around the site of meteorological data collection ensuring that the weather that was recorded closely matched that actually experienced by the infants. Nearly all SIDS deaths occurred within a 20km radius of the meteorological station. In comparison some temperature studies have extrapolated measures from one site homogenously over an entire region or country (Julious, 1997; Jones et al., 1994; Ponsonby et al., 1992; McGlashan & Grice, 1983). Such assumptions can lead to a substantial degree of measurement error, particularly if adopted for variables such as wind.

The strengths of the study relating to the statistical analyses include:

1. This is the first study to statistically identify points of change in the temporal sequence of SIDS counts. These points of change correspond to structural shifts in the distribution underlying the observed process. To successfully model the relationship between the incidence of SIDS and climate these characteristics of the chronological SIDS profile were incorporated into the statistical analyses.

2. Seasonality has been the only consistent finding in studies relating SIDS to climate, yet selection of variables to describe the seasonal SIDS profile appear somewhat arbitrary. This study examined a variety of seasonality measures seen in the literature, along with newly derived variables based on varying climatic measures, to identify the best variable to describe the underlying seasonal fluctuations evident in the SIDS counts in Canterbury.

3. A systematic approach was taken to achieve careful, comprehensive, statistical analyses relating the incidence of SIDS in Canterbury to a climatic profile. Analyses were performed at both a daily and monthly level, with significant climatic effects reported over and above the seasonality inherent in the SIDS profile. Methods of analyses ranged from standard statistical generalised linear models (logistic regression and Poisson regression), to the creation of climatic components to include as regressors in the model, and extensions to the Poisson regression model to include varying mixture models, autoregressive latent processes, and non-linear climatic functionals.

The data set utilised in this thesis has been one of the main strengths of the study, yet was restricted to meteorological variables measured at one (central) station. The environment directly related to a SIDS infant is the indoor temperature, humidity, and local pollutants. A strong association between indoor and outdoor temperature was shown by Schluter et al. (2000), which is a direct result of the typically poorly insulated houses in Christchurch, in which it is unusual to have central heating. This, in part, justifies the use of outdoor temperature as a proxy for the temperature in the immediate environment of the infant. It has been hypothesised that some of the climatic relationships found with the incidence of SIDS are not directly causative as such, but that varying climatic patterns relate to changes in the infants daily routine, through their caregivers (Macey et al., 2000; Auliciems & Barnes, 1987). For example, Auliciems & Barnes (1987) showed that average visibility increased prior to SIDS in both summer and winter cases in Brisbane. Auliciems & Barnes (1987) suggest that this finding may be a result of increased outdoor exposure of infants, or the clear weather changing the parental behaviour.

At present, there is a lot of research that is generally focused on examining the effect of pollution measures on various morbidity and mortality outcomes (for example Goldberg et al. (2003); Campbell & Tobias (2000); Schwartz et al. (1996)). A concomitant regional measure of pollution in Canterbury was not available for the period of the study. Visibility was considered as a proxy, following the work of Knöbel et al. (1995), but poorly predicted pollution over a period of two winter seasons.

A smoke free environment has been shown to greatly reduce the risk of SIDS (Pollack, 2001; Wisborg et al., 2000) indicating an effect of pollution localised in an infant's immediate environment. Recent studies have found associations between environmental pollution (particulate matter and nitrate levels in water) and the incidence of SIDS (Lipfert et al., 2000; George et al., 2001). Therefore examining the relationship between general measures of air pollution and the incidence of SIDS should be considered in future SIDS research in Canterbury, especially as Christchurch has high pollution levels in winter (Fisher & Taylor, 1999).

Information on known risk factors of SIDS, such as smoking, breastfeeding, or economic status was unavailable for the complete set of SIDS in the study population. Although, ideally a statistical analysis would adjust for these variables, this should not detract from the results relating to climate presented in this study. Ninety percent of SIDS deaths occurred within a twenty kilometre radius of the meteorological recordings, and the general Christchurch area has a homogeneous climatic pattern. Therefore, assuming a uniform effect of climate across most SIDS cases was justified — whether a parent breastfed their infant, slept their infant in the supine position, or belonged to a lower socio-economic class, would not have an effect on the weather the infant experienced.

9.5 Future Directions in SIDS Research

Even though SIDS has a long history, research into the epidemiology of SIDS still appears widely in the literature. A general MEDLINE search for "sudden infant death syndrome" revealed 216 articles published in 2002, which were either directly, or indirectly, contributed to the understanding of SIDS. Avenues of research are broad, ranging from: magnesium depletion (Durlach et al., 2002), methyl parathion contamination (Wasley et al., 2002), circulatory failure (Matthews, 2002), and toxic elements in the liver tissue (Lyon et al., 2002); to examining the genetic component (Gordon et al., 2002b), assessing infant physiology and neuronal development (Morgan et al., 2002; Ansari et al., 2002) and quantifying trace elements in the brain (Nishida & Takashima, 2002). Research continues on the role of nicotine exposure (Gordon et al., 2002a; Walsh et al., 2002), and prone sleeping (Moon & Omron, 2002), and the variation of risk factors between ethnicities (Iyasu et al., 2002; Hauck et al., 2002).

SIDS rates are known to vary at an international level (see Section 1.4), and perhaps focusing on these differences may enhance the understanding of the underlying influences on SIDS. The International Child Care Practices study (Nelson & Taylor, 1999) set out to document child care practices in as many different countries and cultures as possible, with the aim of providing baseline child care data and stimulating new hypotheses to explain persisting differences in SIDS rates between countries. Reports have been published from this study detailing international differences in infant sleep position and parental smoking habits (Nelson & Taylor, 2001), and infant sleeping environments (Nelson et al., 2001). The results highlight interesting patterns in child care practices across diverse populations. Nelson et al. (2001) caution that any differences identified should not imply that specific child care practices change the risk of SIDS but suggest that their findings "should help inject caution into the process of developing SIDS prevention campaigns for non-western cultures".

Unlike most causes of death, SIDS is not a specific disease or illness, it is a diagnosis category conveniently used when infants die suddenly and unexpectedly, for no obvious reason. Statisticians have played a vital role in SIDS research. Modifiable risk factors were identified and publicised in the early 1990s, resulting in a dramatic reduction in the number of infant deaths classified as SIDS. It may be that no single cause of SIDS exists, but that different infants die for different reasons. By combining studies, such as the International Child Care Practices Study, with novel statistical approaches, such as those presented in this thesis, differing risk populations and causative factors may be highlighted. SIDS research is a good example of the important role statistics plays in identifying risk factors, and underlying causes, which can lead to a significant reduction in mortality rates.

Appendix A

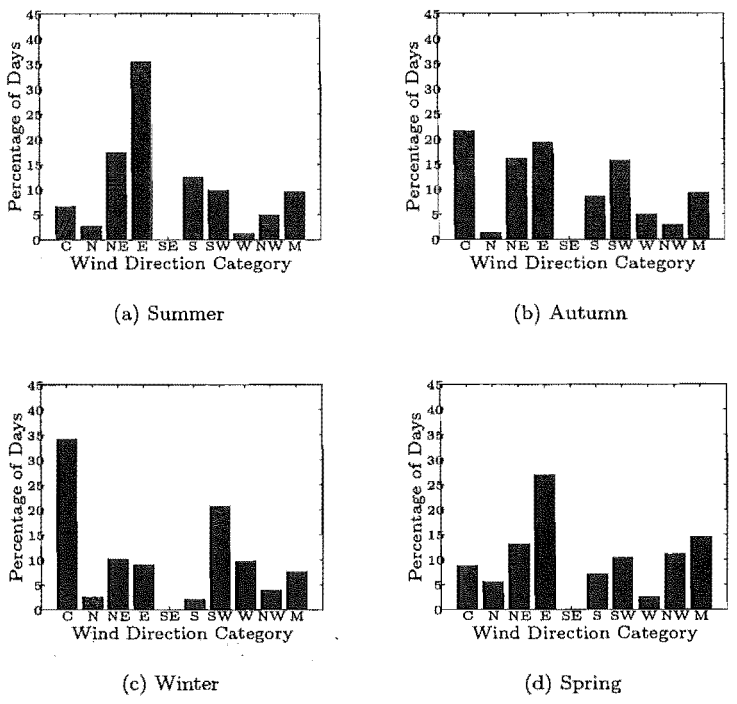
International SIDS Numbers

Country	WHO SIDS rates (year)	Task Force SIDS rates (year)
Argentina	0.22 (1995)	0.56 (1996)
Australia	0.63 (1995)	0.54 (1998)
Austria	0.52 (1998)	0.60 (1998)
Belgium		0.60 (2000)
Cananda	0.40 (1997)	0.45 (1998)
Costa Rica	0.12 (1995)	
Croatia	0.19 (1997)	
Cuba	0.73 (1996)	
Czech Republic	0.08 (1998)	0.90
Denmark	0.36 (1996)	0.30 (1997)
Estonia	0.33 (1998)	
Finland	0.21 (1996)	0.25 (1996)
France	0.56 (1996)	0.49 (1997)
Germany	0.78 (1997)	0.78 (1998)
Greece	0.10 (1997)	0.43 (1991)
Hong Kong		0.10 (1998)
Hungary	0.30 (1998)	0.10 (1998)
Iceland	0.23 (1996)	
Ireland (Republic)	0.57 (1996)	0.90 (1998)
Israel	0.23 (1996)	0.20(1998)
Italy	0.09 (1995)	1.0
Japan	0.37 (1997)	0.30 (1998)
Latvia	0.65 (1998)	
Republic of Moldova	0.13 (1996)	
Netherlands		0.14 (1998)
New Zealand	1.48 (1996)	1.04 (1998)
Norway	0.46 (1995)	0.60 (1998)
Portugal	0.06 (1998)	0.55 (1994-6)
Slovakia		0.14 (1997)
Slovenia	0.22 (1997)	0.47 (1997)
Spain	0.18 (1995)	
Sweden	0.30 (1996)	0.45 (1998)
Switzerland		0.44 (1996)
Russia		0.43 (1992)
United Kingdom (England and Wales)	0.51 (1997)	0.45 (1998)
United Kingdom (Northern Ireland)	0.25 (1997	
United Kingdom (Scotland)	0.70 (1997)	
Ukraine		0.70 (1993)
United States	0.72 (1997)	0.77 (1997)

Table A.1: International rates of SIDS per 1000 live births (WHO = World Health Organisation).

Appendix B

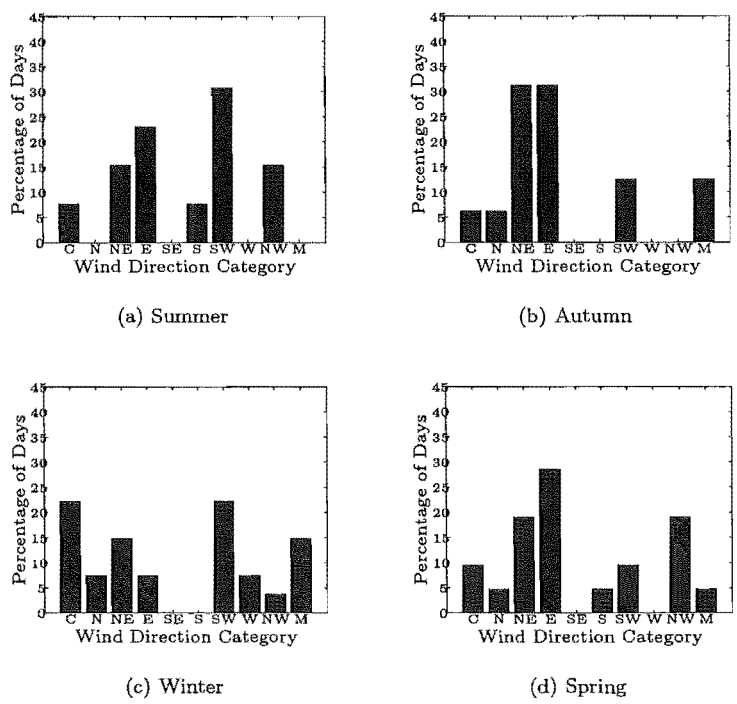
Climatic Variables



(a) Summer (b) Autumn

(c) Winter (d) Spring

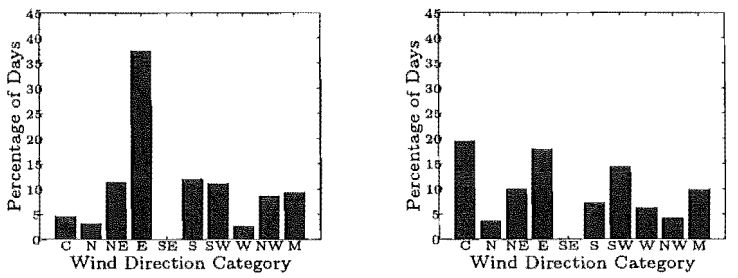
Figure B.1: Percentages of Daily Wind Direction Categories (**No SIDS**, Period 1, *day*0).



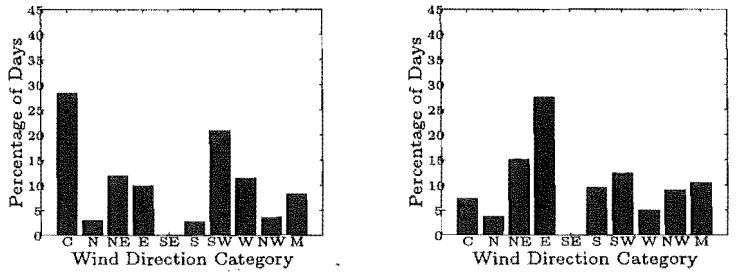
(a) Summer (b) Autumn

(c) Winter (d) Spring

Figure B.2: Percentages of Daily Wind Direction Categories (**SIDS**, Period 1, *day*0).

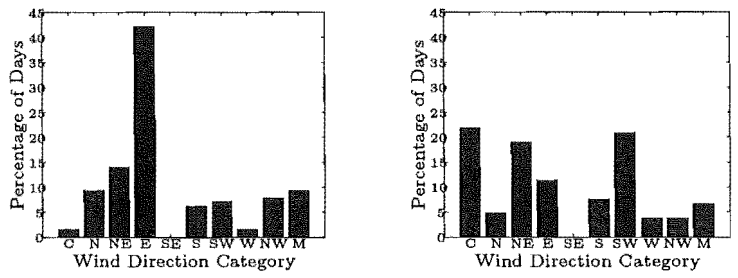


(a) Summer (b) Autumn

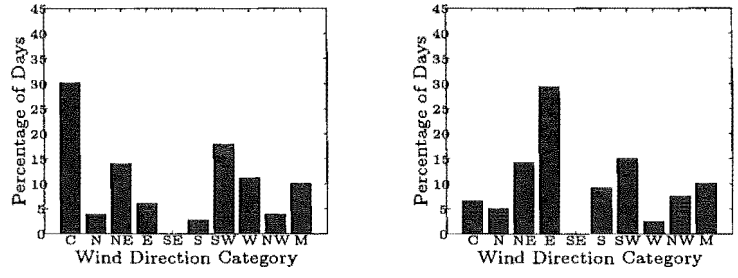


(c) Winter (d) Spring

Figure B.3: Percentages of Daily Wind Direction Categories (**No SIDS**, Period 2, day0).



(a) Summer (b) Autumn



(c) Winter (d) Spring

Figure B.4: Percentages of Daily Wind Direction Categories (**SIDS**, Period 2, day0).

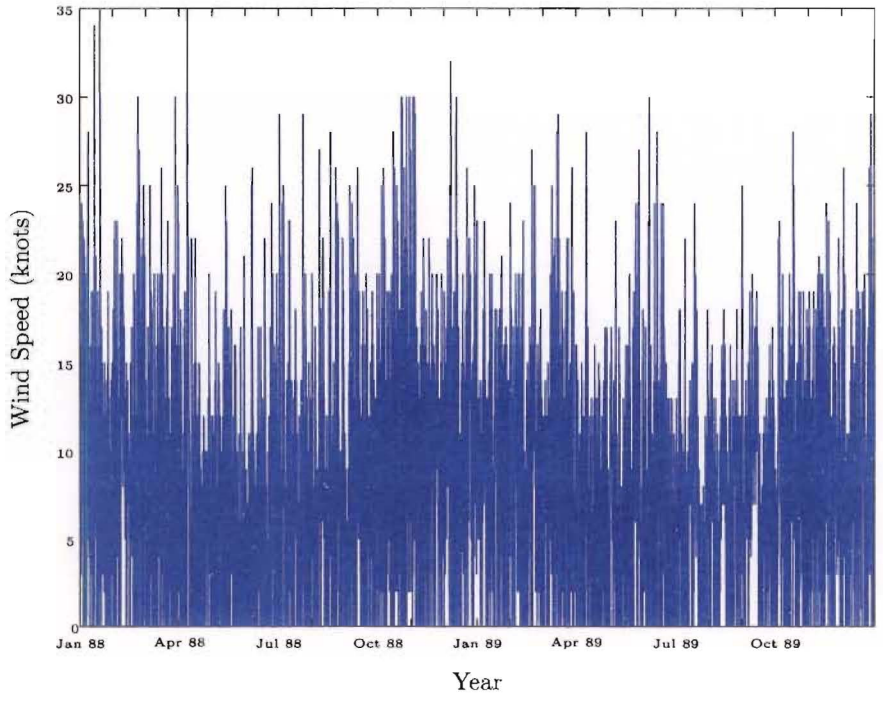
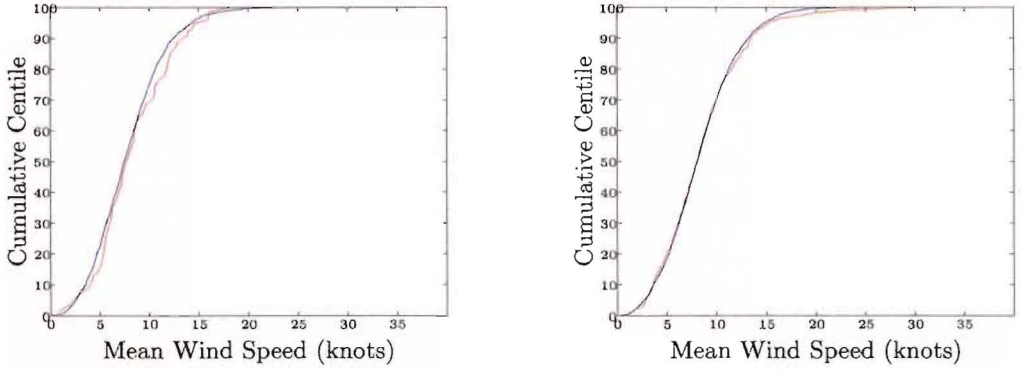


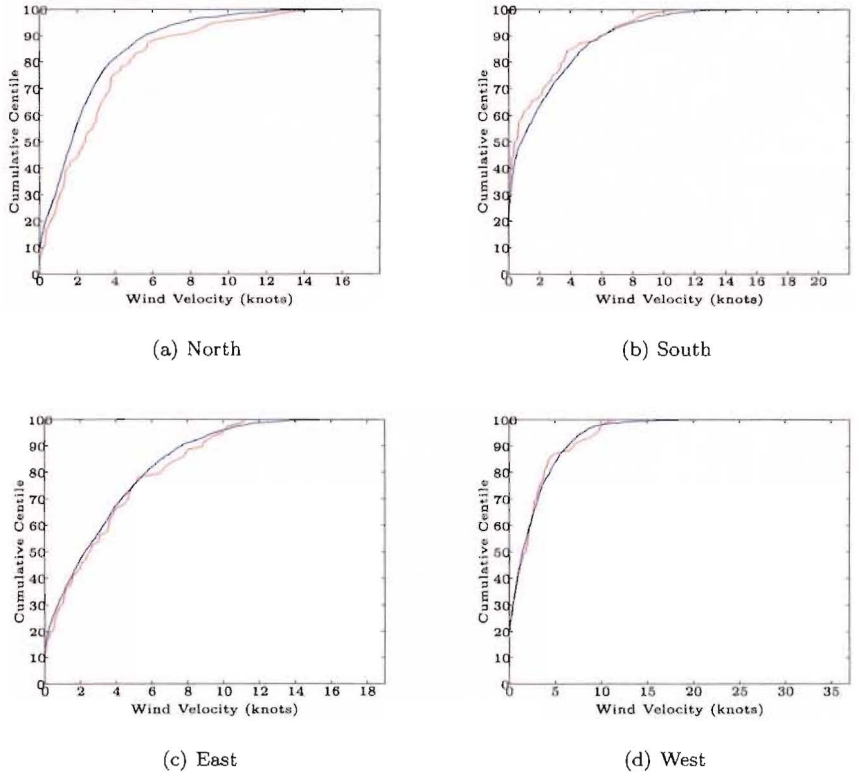
Figure B.5: Hourly wind speed measures for 1988–1989.



(a) Period 1

(b) Period 2

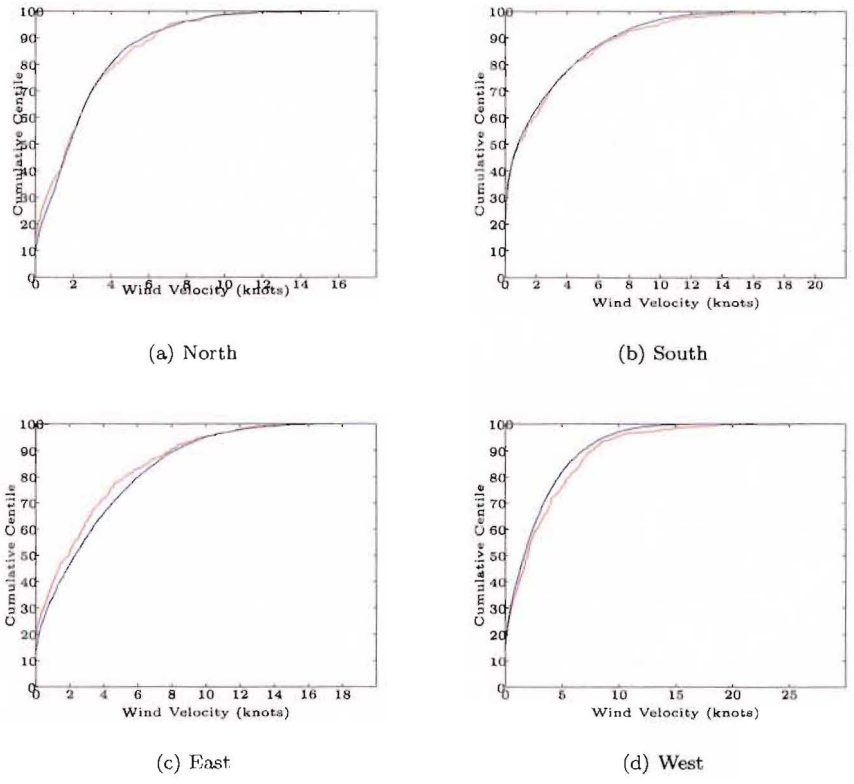
Figure B.6: Cumulative centile distribution for mean daily wind speed (— SIDS days, — non-SIDS days).



(a) North (b) South

(c) East (d) West

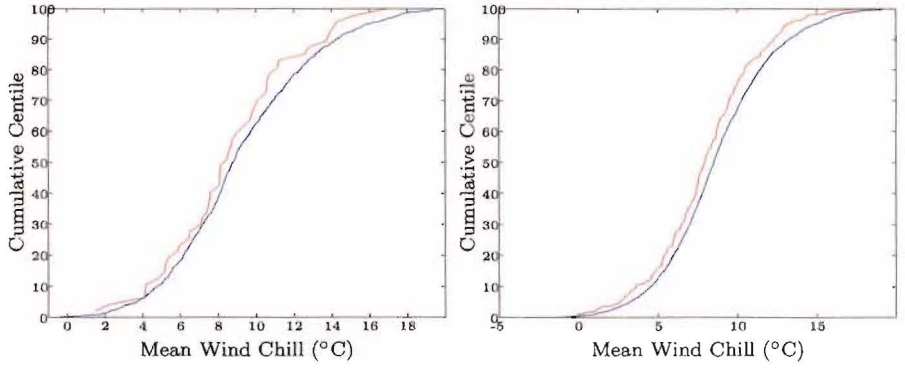
Figure B.7: Cumulative centile distribution for mean daily wind velocity, Period 1 (— SIDS days, — non-SIDS days).



(a) North (b) South

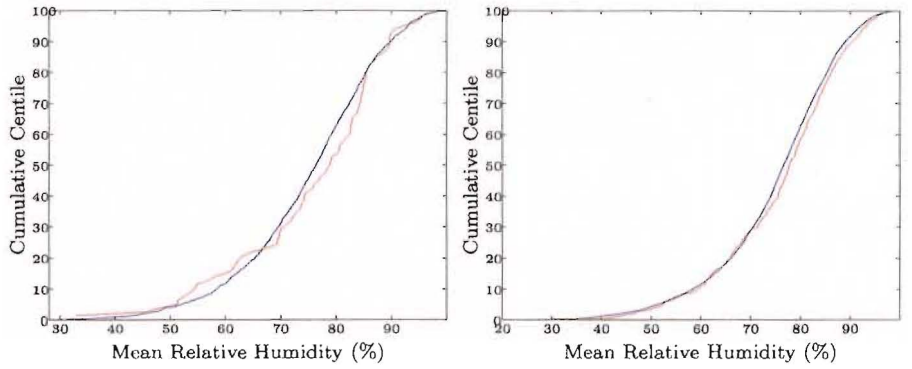
(c) East (d) West

Figure B.8: Cumulative centile distribution for mean daily wind velocity, Period 2 (— SIDS days, — non-SIDS days).



(a) Period 1 (b) Period 2

Figure B.9: Cumulative centile distribution for mean daily wind chill (— SIDS days, — non-SIDS days).



(a) Period 1 (b) Period 2

Figure B.10: Cumulative centile distribution for mean daily humidity (— SIDS days, — non-SIDS days).

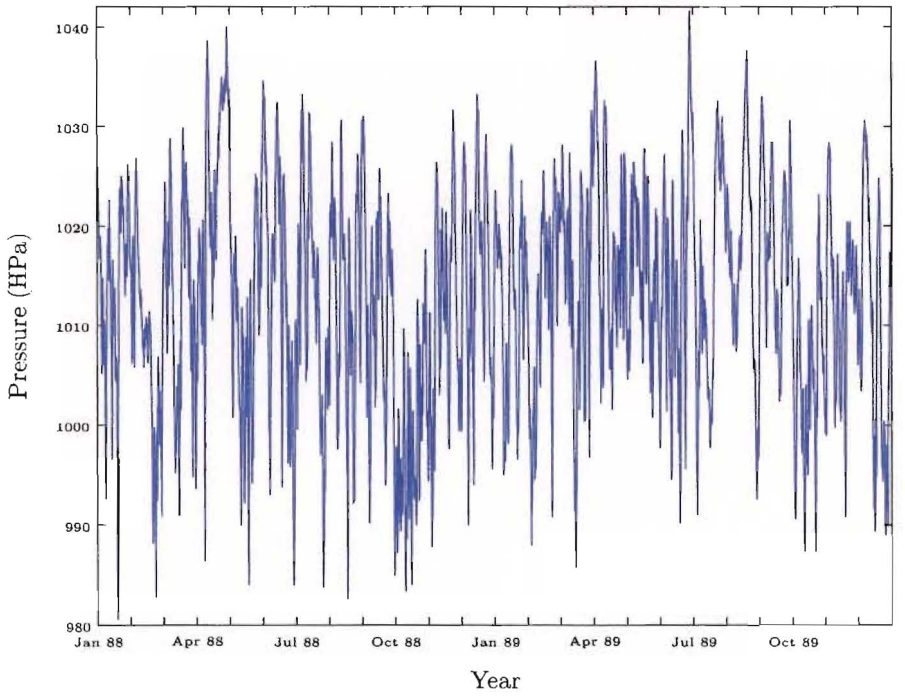
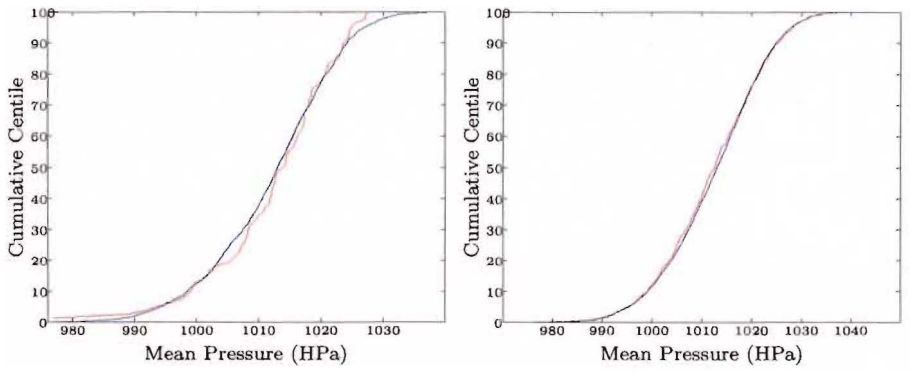
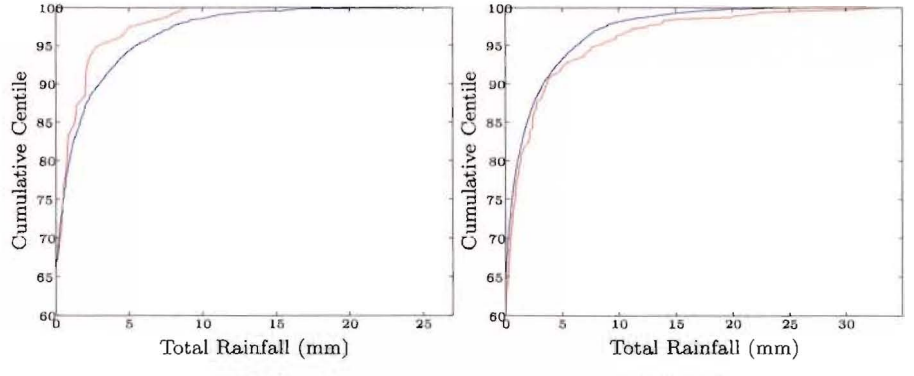


Figure B.11: Hourly pressure measures for 1988—1989.



(a) Period 1 (b) Period 2

Figure B.12: Cumulative centile distribution for mean daily pressure (— SIDS days, — non-SIDS days).



(a) Period 1 (b) Period 2

Figure B.13: Cumulative centile distribution for total daily rainfall (— SIDS days, — non-SIDS days).

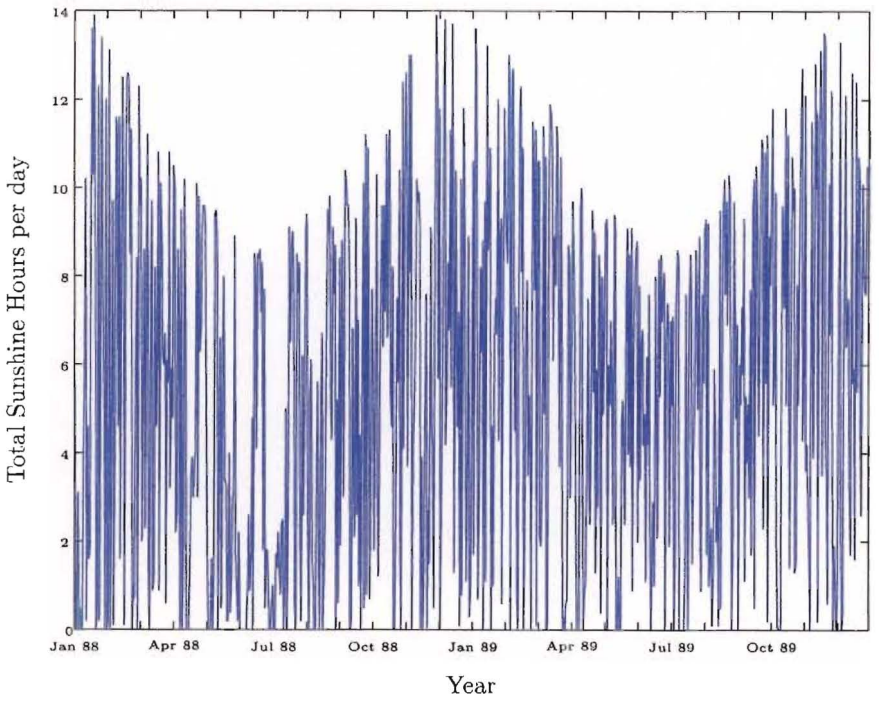
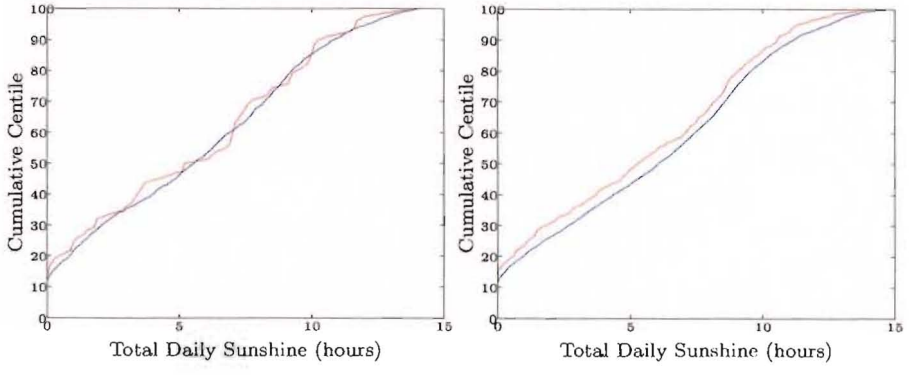


Figure B.14: Daily sunshine measures for 1988–9.



(a) Period 1 (b) Period 2

Figure B.15: Cumulative centile distribution for total daily sunshine (— **SIDS days,** — non-SIDS days).

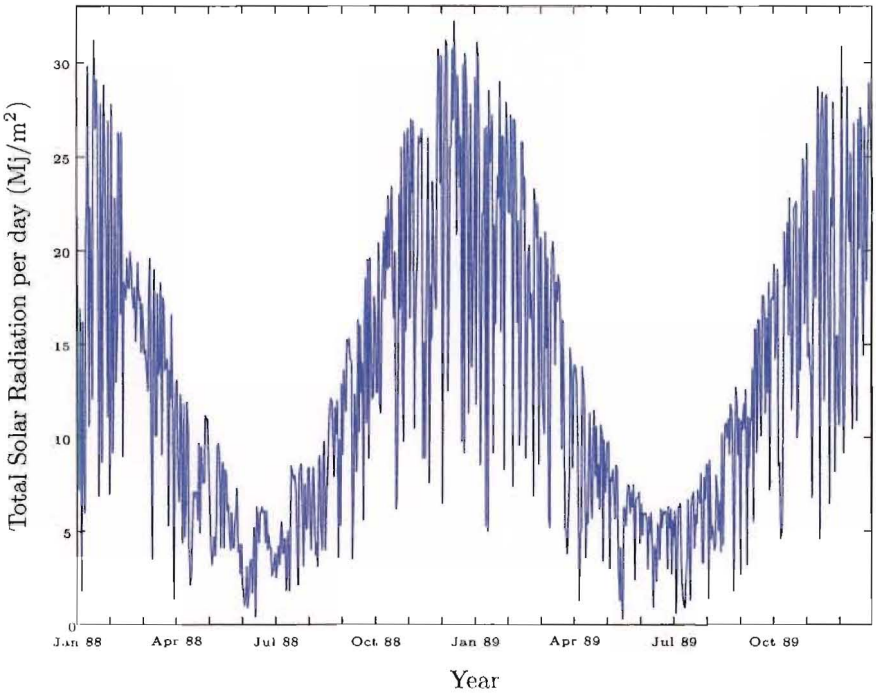
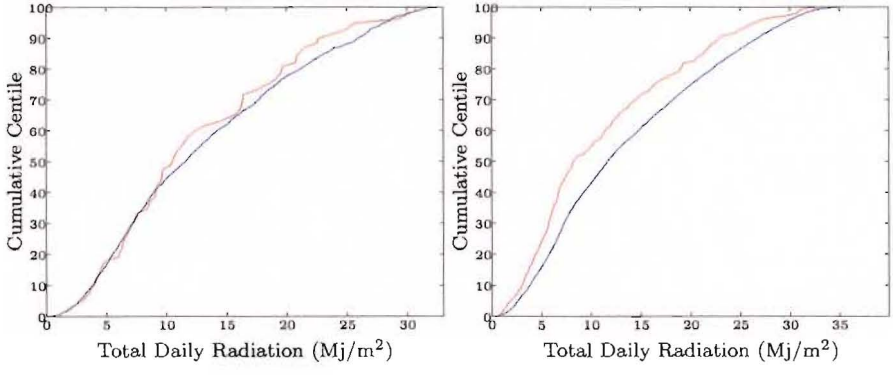


Figure B.16: Daily solar radiation measures for 1988—9.



(a) Period 1 (b) Period 2

Figure B.17: Cumulative centile distribution for total daily solar radiation (— SIDS days, — non-SIDS days).

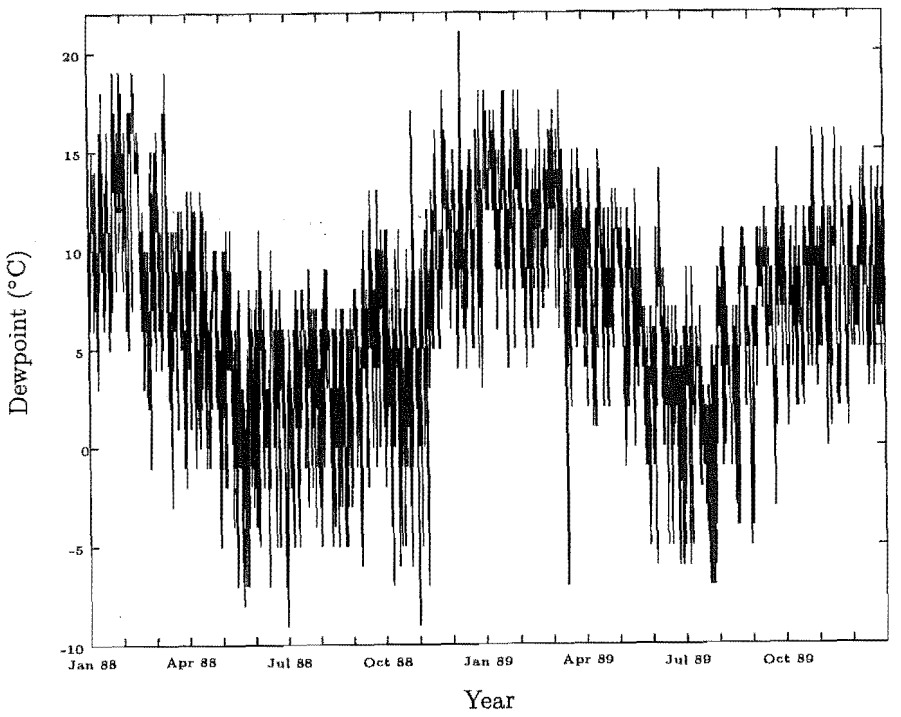


Figure B.18: Hourly dewpoint measures for 1988—1989.

Appendix C

Analysis of Daily Variables

This appendix gives a full list of the climate variables identified in Chapter 3 as being significantly related to the incidence of SIDS at a daily level, over and above any seasonality component. These results are presented in Tables C.1 to C.29, which give parameter estimates, associated odds ratios and goodness-of-fit statistics for each model. Alongside these is a detailed commentary discussing selection between competing variables, parameter interpretations and underlying SIDS risk.

C.1 Period 1 (1968—1972)

C.1.1 Temperature

Results from the logistic regression analysis of daily temperature variables are presented in Table C.1. Only those variables found to be significantly related to SIDS (over and above the baseline seasonality model) appear. The table details parameter estimates, alongside odds ratios and goodness-of-fit statistics. Variables that have been adjusted for seasonality are superscripted with a 'd' (deseasoned), while those that will be retained for further analysis are superscripted with a '$*$'.

Daily summary measures: The variables $Temp_{minday0}$ and $Temp_{meanday0}$, corresponding to adjusted minimum and mean temperature on the day of interest ($day0$) were the only daily summary measures to be related to the rate of SIDS, over and above the baseline seasonality model. The deviance statistics (Equation 3.11) for $Temp_{minday0}$ and $Temp_{meanday0}$ were $G = 6.552$ ($p = 0.010$) and $G = 5.247$ ($p = 0.022$) respectfully, implying that adding either variable to the baseline model significantly improved the log-likelihood. As both these variables are strongly related, and are measuring very similar climatic profiles, only one will be retained

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.056 (0.030)	0.945 (0.891, 1.003)			
$Temp^{d*}_{minday0}$	0.084 (0.038)	1.088 (1.009, 1.173)	627.304	1798.603	4.723
$Temp_{meanMA30}$	-0.057 (0.030)	0.944 (0.890, 1.002)			
$Temp^{d}_{meanday0}$	0.077 (0.041)	1.080 (0.998, 1.170)	687.609	1801.523	4.630
$Temp_{meanMA30}$	-0.061 (0.030)	0.941 (0.887, 0.999)			
$Temp^{*}_{stdday-7}$	0.184 (0.083)	1.201 (1.022, 1.413)	627.343	1808.348	11.077
$Temp_{meanMA30}$	-0.062 (0.030)	0.940 (0.885, 0.997)			
$Temp^{*}_{meanday-6-7}$	0.125 (0.064)	1.133 (1.000, 1.285)	627.304	1808.701	1.875
$Temp_{meanMA30}$	-0.070 (0.031)	0.932 (0.877, 0.991)			
$Temp_{maxday-6-7}$	0.072 (0.041)	1.075 (0.992, 1.164)	629.198	1802.996	12.120
$Temp_{meanMA30}$	-0.066 (0.031)	0.936 (0.882, 0.994)			
$Temp^{*}_{stdday-6-7}$	0.222 (0.109)	1.248 (1.009, 1.544)	628.196	1799.852	2.063
$Temp_{meanMA30}$	-0.054 (0.030)	0.948 (0.893, 1.006)			
$Temp^{*}_{diffday0-14}$	0.085 (0.041)	1.089 (1.004, 1.180)	628.058	1800.420	6.164

Table C.1: Results from logistic regression of Temperature, Period 1 ('d' = deseasoned, '*' = retained for further analysis).

Average Day effects	Estimate (SE) $WindD_{day-1}$	Estimate (SE) $WindD_{day-7}$	Estimate (SE) $WindD_{day-8}$
1	0.008 (0.390)	-0.576 (0.386)	1.504 (0.624)
[Odds ratio (95% CI)]			[4.500 (1.325, 15.281)]
2	-0.151 (0.671)	-0.837 (0.773)	0.069 (1.165)
3	-0.242 (0.430)	-0.589 (0.416)	0.671 (0.685)
4	-0.372 (0.405)	-0.826 (0.399)	0.963 (0.638)
6	-0.320 (0.781)	-0.844 (0.584)	0.716 (0.776)
7	-1.235 (0.550)	-1.187 (0.476)	1.090 (0.650)
[Odds ratio (95% CI)]	[0.291 (0.099, 0.855)]	[0.305 (0.120, 0.776)]	
8	-0.110 (0.564)	-0.657 (0.592)	0.709 (0.834)
9	-1.208 (0.779)	-0.308 (0.505)	1.119 (0.741)
10 (*reference category*)	0	0	0
Deviance	618.742	623.960	620.755
Pearson	1789.471	1786.916	1790.351
Hosmer-Lemeshow	8.136	6.464	4.128

Table C.2: Results from logistic regression of predominante Wind Direction, Period 1 ('d' = deseasoned, '*' = retained for further analysis).

for further analysis. Because of its lower D and P statistics, $Temp_{minday0}$, as compared to $Temp_{meanday0}$, will be retained. These results suggested that the risk of SIDS was greatest when $Temp_{meanMA30}$ was low and $Temp_{minday0}$ high, that is on warmer days in winter. This trend was also identified by Schluter et al. (1998).

Within day effects: Only the standard deviation of the temperature seven days before $day0$ ($Temp_{stdday-7}$) was found to be significantly related to SIDS incidence. The odds ratios for this model indicated a decreased SIDS risk of 0.941 times for each 1°C increase in $Temp_{meanMA30}$, and for $Temp_{stdday-7}$, the risk increased by 1.201. This model suggested that the greatest SIDS risk was in winter, a week after a day where the temperature varied widely.

Between day effects: Three temperature variables corresponding to the difference from six to seven days prior to $day0$ were found to be associated with SIDS, namely $Temp_{meanday-6-7}$, $Temp_{maxday-6-7}$ and $Temp_{stdday-6-7}$. Both $Temp_{meanday-6-7}$ and $Temp_{maxday-6-7}$ are strongly related, and are essentially capturing the same climatic profile. Therefore only $Temp_{meanday-6-7}$, with lower D and HL statistics as compared to $Temp_{maxday-6-7}$, will be retained for further analysis. The model containing $Temp_{meanday-6-7}$ implies that a 1°C increase in $Temp_{meanMA30}$ resulted in SIDS risk decreasing by 0.940, while for $Temp_{meanday-6-7}$, SIDS risk increased 1.133 times. Essentially, this implied an increased SIDS risk in winter where the temperature was higher six days before the day of interest than it was seven days before the day of interest.

Average day effects: No significant associations were found between any of the temperature average day effects and SIDS incidence.

Between average day effects: The only significant relationship with SIDS involved $Temp_{diffday0-14}$, the difference in temperature between the day of interest and the average temperature over the past fortnight. This relationship can be interpreted as an increased SIDS risk in winter on days where the temperature was cooler than it had been over the past two weeks.

C.1.2 Wind Direction

Results from the analysis of wind direction in relation to SIDS incidence appear in Table C.2 and Table C.3. Three lagged variables relating to the predominant wind direction were found to be significantly associated with SIDS, over and above seasonality, as shown in Table C.2. Taking the 'mixed' category (no one direction predominated) as the reference group, SIDS risk decreased 0.291 times when the predominant wind direction on $day-1$ was south-west. A similar south-westerly effect a week ago was also related to decreased SIDS risk, where as, if there was a predominately calm day eight days before $day0$, SIDS risk increased.

The significant relationships between the variables measuring absolute hourly change in wind direc-

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.049 (0.030)	0.952 (0.897, 1.010)			
$WindD^{*}_{maxAHCday-4}$	-0.006 (0.003)	0.994 (0.988, 0.999)	627.217	1798.062	2.309
$Temp_{meanMA30}$	-0.053 (0.030)	0.948 (0.894, 1.006)			
$WindD_{stdAHCday-4}$	-0.024 (0.012)	0.977 (0.954, 1.000)	627.826	1808.688	2.826
$Temp_{meanMA30}$	-0.064 (0.031)	0.938 (0.884, 0.996)			
$WindD^{*}_{maxAHCday-1-2}$	0.005 (0.003)	1.005 (1.000, 1.011)	628.766	1790.918	5.829
$Temp_{meanMA30}$	-0.055 (0.030)	0.946 (0.892, 1.004)			
$WindD^{*}_{meanAHCday-2-3}$	0.015 (0.013)	1.016 (0.990, 1.041)	627.912	1805.081	6.203

Table C.3: Results from logistic regression of variation in daily Wind Direction, Period 1 ('*' = retained for further analysis).

tion and SIDS are presented in Table C.3. Four models are shown, corresponding to $WindD_{maxAHCday-4}$, $WindD_{stdAHCday-4}$, $WindD_{maxAHCday-1-2}$ and $WindD_{meanAHCday-2-3}$. $WindD_{maxAHCday-4}$, with lower values across all three goodness-of-fit statistics compared to $WindD_{stdAHCday-4}$, will be retained a measure of the variability in wind direction four days prior to the day of interest in further analysis. An increase in $WindD_{maxAHCday-4}$ corresponded to a decrease in SIDS risk, whereas an increase in $WindD_{maxAHCday-1-2}$ or $WindD_{meanAHCday-2-3}$ corresponded to an increase in the risk of SIDS.

C.1.3 Wind Speed

Wind speed was not significantly associated with seasonality. However, the addition of most wind speed variables to the baseline model of intercept and seasonality as measured by $Temp_{meanMA30}$, resulted in $Temp_{meanMA30}$ no longer being significant. The models corresponding to significant wind speed variables are presented in Table C.4.

Daily summary measures: Lagged daily measures of both mean and maximum wind speed were identified as being significantly related to SIDS incidence: mean and maximum wind speed one and two days before *day*0 ($WindS_{meanday-1}$, $WindS_{meanday-2}$, $WindS_{maxday-1}$, and $WindS_{maxday-2}$,) and mean wind speed a week before *day*0 ($WindS_{maxday-7}$,). Of the two variables measuring wind speed one and two days before *day*0, the mean wind speed will be retained as it has lower values for two of the three goodness-of-fit statistics as compared to maximum wind speed. All three mean wind speed measures inferred a decreased SIDS risk with increased wind speed.

Within day effects: Seven within day effects were found to be significantly related to SIDS, for two, eight and fourteen day lags, as detailed in Table C.4. Comparison of the D, P, and HL statistics showed that $WindS_{rangeday-2}$, $WindS_{maxAHCday-8}$, and $WindS_{stdday-14}$ scored best on two out of the three statistics when compared to other within day effects over the same lagged time frame. Therefore these variables will be retained for further analysis. As with the day effects, the three retained within day effects corresponded to a decreased SIDS risk on days with increased variability in wind speed.

Between day effects: A variety of between day effects were found to be related significantly to SIDS incidence, ranging from $WindS_{meanday-1-2}$ to $WindS_{stdday-7-8}$. Those that will be retained for further analysis are $WindS_{meanday-1-2}$, $WindS_{maxAHCday-7-8}$ and $WindS_{meanAHCday-4-5}$, all of which significantly improve the log-likelihood over the baseline model of seasonality only (G ranging from 4.225 ($p = 0.040$) to 9.993 ($p = 0.001$)). Both $WindS_{meanday-1-2}$ and $WindS_{maxAHCday-7-8}$ related to a decreased SIDS risk. This implied that SIDS risk decreased when the mean wind speed was greater on $day - 1$, than it was on $day - 2$, and similarly, SIDS risk decreases when the maximum absolute hourly change in wind speed seven days prior to the day of interest was greater than it was eight days prior to the day of interest. In contrast SIDS risk increased when the mean absolute hourly change was less four days ago than five days ago.

Average day effects: Of the eight average day effects examined in relation to SIDS, six were significant. As one of the main aims of this section of analysis is data reduction only two are retained, namely

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.983, 1.109)			
$WindS^{*}_{meanday-1}$	-0.076 (0.035)	0.927 (0.865, 0.993)	628.914	1785.299	8.798
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.983, 1.109)			
$WindS^{*}_{meanday-2}$	-0.077 (0.035)	0.926 (0.864, 0.992)	628.735	1799.846	4.035
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.983, 1.109)			
$WindS^{*}_{meanday-7}$	-0.092 (0.036)	0.912 (0.851, 0.978)	626.750	1809.393	10.355
$Temp_{meanMA30}$	0.042 (0.031)	1.043 (0.981, 1.108)			
$WindS_{maxday-1}$	-0.044 (0.022)	0.957 (0.918, 0.999)	629.455	1773.705	10.781
$Temp_{meanMA30}$	0.042 (0.031)	1.042 (0.981, 1.108)			
$WindS_{maxday-2}$	-0.062 (0.022)	0.940 (0.900, 0.982)	625.465	1800.732	7.879
$Temp_{meanMA30}$	0.042 (0.031)	1.043 (0.982, 1.108)			
$WindS_{stdday-2}$	-0.160 (0.078)	0.852 (0.731, 0.993)	629.332	1823.612	4.4512
$Temp_{meanMA30}$	0.043 (0.031)	1.043 (0.982, 1.109)			
$WindS^{*}_{stdday-14}$	-0.794 (0.080)	0.824 (0.704, 0.964)	627.507	1813.284	7.537
$Temp_{meanMA30}$	0.042 (0.031)	1.042 (0.981, 1.107)			
$WindS^{*}_{rangeday-2}$	-0.069 (0.024)	0.933 (0.890, 0.978)	624.649	1820.924	5.413
$Temp_{meanMA30}$	0.040 (0.031)	1.041 (0.980, 1.106)			
$WindS^{*}_{maxAHCday-8}$	-0.844 (0.039)	0.920 (0.852, 0.993)	628.424	1788.203	2.602
$Temp_{meanMA30}$	0.039 (0.031)	1.040 (0.979, 1.105)			
$WindS_{meanAHCday-2}$	-0.308 (0.148)	0.735 (0.549, 0.983)	629.073	1811.657	12.295
$Temp_{meanMA30}$	0.039 (0.031)	1.040 (0.979, 1.105)			
$WindS_{meanAHCday-8}$	-0.324 (0.150)	0.723 (0.539, 0.971)	627.8.650	1806.081	10.102
$Temp_{meanMA30}$	0.401 (0.031)	1.041 (0.980, 1.106)			
$WindS_{stdAHCday-8}$	-0.335 (0.166)	0.716 (0.516, 0.991)	629.249	1790.062	5.559
$Temp_{meanMA30}$	0.040 (0.031)	1.040 (0.979, 1.105)			
$WindS^{*}_{meanday-1-2}$	-0.105 (0.054)	0.900 (0.810, 1.001)	629.631	1794.853	11.040
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.982, 1.109)			
$WindS_{stdday-4-5}$	0.165 (0.080)	1.075 (1.009, 1.378)	629.887	1781.318	16.955
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.982, 1.109)			
$WindS^{*}_{maxAHCday-7-8}$	-0.140 (0.050)	0.870 (0.788, 0.960)	623.863	1749.900	16.138
$Temp_{meanMA30}$	0.046 (0.031)	1.047 (0.985, 1.113)			
$WindS^{*}_{meanAHCday-4-5}$	0.310 (0.141)	1.364 (1.034, 1.798)	629.452	1799.473	5.022
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.982, 1.110)			
$WindS_{stdAHCday-7-8}$	-0.512 (0.109)	0.599 (0.393, 0.915)	626.952	1768.827	16.809
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.983, 1.109)			
$WindS_{meanday-1to-2}$	-0.110 (0.042)	0.896 (0.823, 0.972)	626.604	1793.276	5.387
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.982, 1.109)			
$WindS^{*}_{meanday-1to-3}$	-0.115 (0.046)	0.892 (0.814, 0.977)	627.561	1780.607	2.084
$Temp_{meanMA30}$	0.043 (0.031)	1.043 (0.982, 1.108)			
$WindS_{meanday-1to-4}$	-0.106 (0.050)	0.899 (0.815, 0.993)	629.318	1780.051	11.099
$Temp_{meanMA30}$	0.042 (0.031)	1.043 (0.982, 1.108)			
$WindS_{meanday-1to-7}$	-0.130 (0.059)	0.878 (0.783, 0.986)	628.949	1785.091	5.581
$Temp_{meanMA30}$	0.042 (0.031)	1.043 (0.982, 1.108)			
$WindS_{meanday-1to-8}$	-0.147 (0.061)	0.863 (0.766, 0.973)	627.953	1783.595	3.014
$Temp_{meanMA30}$	0.042 (0.031)	1.043 (0.982, 1.107)			
$WindS^{*}_{meanday-1to-14}$	-0.147 (0.069)	0.864 (0.754, 0.989)	629.354	1791.668	5.130
$Temp_{meanMA30}$	0.044 (0.031)	1.045 (0.983, 1.111)			
$WindS^{*}_{diffday0-2}$	0.085 (0.028)	1.089 (1.031, 1.151)	624.941	1774.572	6.221
$Temp_{meanMA30}$	0.044 (0.031)	1.045 (0.983, 1.110)			
$WindS_{diffday0-3}$	0.078 (0.028)	1.080 (1.022, 1.141)	626.713	1780.575	6.866
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.983, 1.110)			
$WindS_{diffday0-4}$	0.069 (0.029)	1.072 (1.013, 1.133)	628.298	1784.217	5.675
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.983, 1.109)			
$WindS_{diffday0-5}$	0.061 (0.029)	1.063 (1.005, 1.125)	629.586	1789.083	4.421
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.983, 1.109)			
$WindS_{diffday0-6}$	0.061 (0.029)	1.063 (1.004, 1.126)	629.707	1792.369	6.024
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.983, 1.110)			
$WindS_{diffday0-7}$	0.068 (0.029)	1.070 (1.011, 1.132)	628.742	1789.851	12.324
$Temp_{meanMA30}$	0.0043(0.031)	1.044 (0.983, 1.110)			
$WindS_{diffday0-8}$	0.070 (0.029)	1.072 (1.013, 1.134)	628.413	1787.973	5.719
$Temp_{meanMA30}$	0.043 (0.031)	1.044 (0.982, 1.109)			
$WindS^{*}_{diffday0-14}$	0.063 (0.029)	1.065 (1.006, 1.127)	629.569	1797.998	11.677

Table C.4: Results from logistic regression of Wind Speed, Period 1 ('*' = retained for further analysis).

$WindS_{meanday-1to-3}$ and $WindS_{meanday-1to-14}$. The first variable was chosen as it scored best on the goodness-of-fit statistics in comparison to the other average day effects. The second variable was retained as it essentially captures all the information about mean wind speed over the two weeks before the day of interest, incorporating information found in the other variables. Both variables showed a decreased SIDS risk with increased average wind speed over the few days prior to $day0$.

Between average day effects: All variables examined in this category were found to be significantly related to SIDS incidence. Using the same method as for the average day effects to reduce the number of variables to be considered in further analysis, only two were retained; namely $WindS_{diffday0to-2}$ with the best overall D, P and HL statistics, and $WindS_{diffday0-14}$ which captures the most information. Both variables related to an increased risk of SIDS when the wind speed on the day of interest was greater than it had been in the past.

C.1.4 Wind Velocity — North

As with wind speed, the northern component of wind velocity was not associated with $Temp_{meanMA30}$, that is, there was no apparent underlying seasonal pattern in the variable. The results for the significant models found via logistic regression are presented in Table C.5.

Daily summary measures: The three daily summary measures corresponding to the northerly component of wind velocity that were found to be significantly associated with SIDS were $North_{meanday0}$, $North_{maxday0}$ and $North_{maxday-8}$. The goodness-of-fit statistics for $North_{maxday0}$ were consistently lower than those of $North_{meanday0}$, and thus of these two similar variables $North_{maxday0}$ is retained for further analysis. Interpretation of the odds ratios showed that for a 1°C increase in $Temp_{meanMA30}$, SIDS risk decreased 0.935 times, with a 1 knot increase in $North_{maxday0}$ relating to a 1.051 times increase in SIDS risk. Adding $North_{maxday-8}$ to the baseline seasonality model results in $Temp_{meanMA30}$ no longer being significantly associated with SIDS ($p = 0.164$), although $North_{maxday-8}$ is retained and the underlying relationship between it and SIDS is examined in further analyses.

Within day effects: All the variables measuring fluctuations in northerly wind velocity within the day of interest ($day0$) and eight days prior ($day - 8$), were found to be significantly associated with SIDS. Of the five variables, $North_{stdday0}$ will be retained due to it having lower values of at least two of the three goodness-of-fit statistics as compared to the other variables. Similarly, of the five variables measuring the eight day lag in within day effects, $North_{stdday-8}$ is retained. The odds ratio for the first model indicated a decreased SIDS risk of 0.935 times for each 1°C increase in $Temp_{meanMA30}$, and for $North_{stdday0}$ the risk increased by 1.160 times with every knot increase in northerly wind velocity. This model suggested that the highest SIDS risk was in winter, on days where the wind velocity from the northerly direction varied widely. With the second model, the addition of $North_{stdday-8}$ to the baseline model resulted in $Temp_{meanMA30}$ no longer being significant, although $North_{stdday-8}$ was.

Between day effects: Nine between day effects were found to be related to SIDS incidence. Those that will be retained for further analysis are $North_{minday0-7}$, $North_{maxday-3-4}$, $North_{maxday0-6}$, $North_{maxAHCday-3-4}$, $North_{maxAHCday-7-8}$, $North_{rangeday0-5}$ and $North_{rangeday0-6}$. The variables that were associated with increased SIDS risk are $North_{minday0-7}$, $North_{maxday0-6}$, $North_{rangeday0-5}$ and $North_{rangeday0-6}$, while $North_{maxday-3-4}$, $North_{maxAHCday-3-4}$ and $North_{maxAHCday-7-8}$ were associated with decreased SIDS risk. Odds ratios for these models are given in Table C.5.

Average day effects: As with temperature, no significant associations between the northerly component of wind velocity average day effects and the incidence of SIDS were found.

Between average day effects: All the variables measuring the difference between north wind velocity on the day of interest and past averages were significantly related to SIDS rates. Two are retained for

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.065 (0.031)	0.937 (0.883, 0.996)			
$North_{meanday0}$	0.100 (0.037)	0.927 (1.029, 1.188)	629.616	1809.048	9.775
$Temp_{meanMA30}$	-0.067 (0.031)	0.935 (0.880, 0.993)			
$North^{*}_{maxday0}$	0.050 (0.017)	1.051 (1.016, 1.087)	625.685	1807.303	13.803
$Temp_{meanMA30}$	-0.044 (0.031)	0.957 (0.901, 1.016)			
$North_{maxday-8}$	-0.047 (0.024)	0.954 (0.910, 1.001)	628.022	1792.013	10.137
$Temp_{meanMA30}$	-0.067 (0.031)	0.935 (0.880, 0.994)			
$North^{*}_{stdday0}$	0.148 (0.058)	1.160 (1.034, 1.300)	626.284	1805.937	3.522
$Temp_{meanMA30}$	-0.042 (0.031)	0.959 (0.903, 1.018)			
$North^{*}_{stdday-8}$	-0.174 (0.081)	0.841 (0.717, 0.985)	627.034	1790.014	7.759
$Temp_{meanMA30}$	-0.067 (0.031)	0.935 (0.880, 0.993)			
$North_{rangeday0}$	0.051 (0.018)	1.052 (1.016, 1.089)	624.681	1807.062	6.019
$Temp_{meanMA30}$	-0.044 (0.031)	0.957 (0.901, 1.017)			
$North_{rangeday-8}$	-0.049 (0.025)	0.952 (0.907, 0.999)	627.754	1790.899	9.814
$Temp_{meanMA30}$	-0.065 (0.031)	0.937 (0.882, 0.995)			
$North_{maxAHCday0}$	0.063 (0.026)	1.065 (1.012, 1.121)	626.910	1799.452	4.542
$Temp_{meanMA30}$	-0.044 (0.031)	0.957 (0.901, 1.016)			
$North_{maxAHCday-8}$	-0.078 (0.036)	0.925 (0.862, 0.993)	627.040	1790.183	7.982
$Temp_{meanMA30}$	-0.063 (0.031)	0.939 (0.884, 0.997)			
$North_{meanAHCday0}$	0.197 (0.099)	1.217 (1.002, 1.478)	628.562	1801.254	16.792
$Temp_{meanMA30}$	-0.044 (0.031)	0.957 (0.901, 1.016)			
$North_{meanAHCday-8}$	-0.283 (0.137)	0.753 (0.576, 0.986)	627.376	1798.546	9.311
$Temp_{meanMA30}$	-0.064 (0.031)	0.938 (0.883, 0.996)			
$North_{stdAHCday0}$	0.196 (0.083)	1.217 (0.994, 1.490)	628.803	1801.933	5.280
$Temp_{meanMA30}$	-0.045 (0.031)	0.956 (0.900, 1.015)			
$North_{stdAHCday-8}$	-0.250 (0.131)	0.779 (0.602, 1.007)	628.257	1789.468	12.041
$Temp_{meanMA30}$	-0.060 (0.030)	0.942 (0.888, 1.000)			
$North^{*}_{minday0-7}$	0.193 (0.095)	1.213 (1.007, 1.460)	629.000	1805.756	12.703
$Temp_{meanMA30}$	-0.058 (0.030)	0.944 (0.890, 1.001)			
$North^{*}_{maxday-3-4}$	-0.073 (0.034)	0.929 (0.869, 0.993)	626.661	1826.150	8.914
$Temp_{meanMA30}$	-0.056 (0.031)	0.946 (0.891, 1.004)			
$North^{*}_{maxday0-5}$	0.037 (0.019)	1.133 (1.000, 1.077)	628.759	1798.988	13.098
$Temp_{meanMA30}$	-0.056 (0.031)	0.946 (0.891, 1.004)			
$North^{*}_{maxday0-6}$	0.039 (0.020)	1.040 (1.000, 1.082)	628.656	1809.079	4.158
$Temp_{meanMA30}$	-0.058 (0.030)	0.944 (0.890, 1.001)			
$North_{rangeday-3-4}$	-0.071 (0.034)	0.932 (0.872, 0.996)	627.093	1823.412	12.017
$Temp_{meanMA30}$	-0.056 (0.031)	0.946 (0.891, 1.004)			
$North^{*}_{rangeday0-5}$	0.038 (0.019)	1.039 (1.000, 1.079)	628.669	1801.128	18.417
$Temp_{meanMA30}$	-0.055 (0.031)	0.946 (0.891, 1.004)			
$North^{*}_{rangeday0-6}$	0.043 (0.020)	1.043 (1.003, 1.086)	628.157	1808.054	6.965
$Temp_{meanMA30}$	-0.059 (0.030)	0.943 (0.889, 1.000)			
$North^{*}_{maxAHCday-3-4}$	-0.090 (0.047)	0.914 (0.833, 1.003)	628.072	1775.995	5.867
$Temp_{meanMA30}$	-0.059 (0.030)	0.943 (0.889, 1.000)			
$North^{*}_{maxAHCday-7-8}$	-0.102 (0.049)	0.903 (0.821, 0.994)	627.096	1794.837	3.191
$Temp_{meanMA30}$	-0.056 (0.030)	0.945 (0.891, 1.003)			
$North_{maxAHCday0-6}$	0.056 (0.029)	1.058 (0.999, 1.120)	628.851	1806.447	7.660
$Temp_{meanMA30}$	-0.056 (0.030)	0.946 (0.891, 1.004)			
$North_{diffday0-2}$	0.115 (0.039)	1.122 (1.040, 1.211)	623.798	1815.261	5.325
$Temp_{meanMA30}$	-0.055 (0.030)	0.946 (0.892, 1.004)			
$North_{diffday0-3}$	0.112 (0.038)	1.119 (1.039, 1.205)	624.062	1813.990	16.011
$Temp_{meanMA30}$	-0.055 (0.030)	0.947 (0.892, 1.005)			
$North^{*}_{diffday0-4}$	0.114 (0.037)	1.121 (1.042, 1.205)	623.662	1811.625	4.350
$Temp_{meanMA30}$	-0.055 (0.030)	0.947 (0.892, 1.005)			
$North_{diffday0-5}$	0.110 (0.037)	1.117 (1.0038 1.201)	624.231	1808.037	6.449
$Temp_{meanMA30}$	-0.055 (0.030)	0.947 (0.892, 1.005)			
$North_{diffday0-6}$	0.111 (0.038)	1.118 (1.038, 1.203)	624.379	1808.585	7.606
$Temp_{meanMA30}$	-0.055 (0.030)	0.947 (0.892, 1.005)			
$North_{diffday0-7}$	0.111 (0.038)	1.117 (1.037, 1.203)	624.601	1806.626	6.325
$Temp_{meanMA30}$	-0.054 (0.030)	0.947 (0.892, 1.005)			
$North_{diffday0-8}$	0.115 (0.037)	1.122 (1.043, 1.207)	623.796	1802.218	9.561
$Temp_{meanMA30}$	-0.054 (0.030)	0.948 (0.893, 1.006)			
$North_{diffday0-14}$	0.112 (0.037)	1.119 (1.041, 1.203)	624.015	1798.420	14.219

Table C.5: Results from logistic regression of Wind Velocity (North), Period 1 ('*' = retained for further analysis).

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.057 (0.030)	0.945 (0.890, 1.003)			
$South^{*}_{meanday-7}$	-0.122 (0.056)	0.927 (0.793, 0.988)	626.492	1784.340	4.864
$Temp_{meanMA30}$	-0.055 (0.030)	0.947 (0.892, 1.004)			
$South_{maxday-7}$	-0.037 (0.019)	0.964 (0.928, 1.000)	628.069	1803.283	8.146
$Temp_{meanMA30}$	-0.055 (0.030)	0.947 (0.892, 1.004)			
$South^{*}_{rangeday-7}$	-0.037 (0.019)	0.964 (0.928, 1.001)	628.273	1804.867	8.319
$Temp_{meanMA30}$	-0.061 (0.031)	0.941 (0.886, 0.998)			
$South^{*}_{maxday0-5}$	0.045 (0.018)	1.046 (1.010, 1.083)	626.169	1831.425	14.209
$Temp_{meanMA30}$	-0.062 (0.031)	0.940 (0.886, 0.998)			
$South^{*}_{stdday0-5}$	0.126 (0.055)	1.134 (1.018, 1.264)	627.196	1830.642	8.087
$Temp_{meanMA30}$	-0.062 (0.031)	0.940 (0.886, 0.998)			
$South_{rangeday0-5}$	0.045 (0.018)	1.046 (1.009, 1.083)	626.375	1830.896	16.505

Table C.6: Results from logistic regression of Wind Velocity (South), Period 1 ('*' = retained for further analysis).

further analysis, namely $North_{diffday0to-4}$ and $North_{diffday0-14}$, the first variable as it rates best over all the between average day effects in terms of the goodness-of-fit statistics, while the second was chosen as it captures the information about the northerly wind velocity over the two weeks prior to the day of interest, essentially incorporating the information measured by the other variables. Both $North_{diffday0to-4}$ and $North_{diffday0-14}$ were associated with increased SIDS risk, implying that the greatest risk of SIDS occurred in winter on days where the northerly component of wind velocity was stronger than it had been in the past few days.

C.1.5 Wind Velocity — South

Like wind speed and northerly wind velocity, the southern component of wind velocity was not significantly associated with $Temp_{meanMA30}$, the seasonal measure. Table C.6 presents the models that found a southerly wind velocity measure significant, over and above the seasonal component.

Daily summary measures: The mean and maximum southern wind velocity seven days before *day*0 were both found to be significantly related to SIDS incidence. With a lower value on all three goodness-of-fit statistics, $South_{meanday-7}$ is retained over $South_{maxday-7}$ for further examination of its relationship with SIDS. The results of this model indicate that a 1°C increase in $Temp_{meanMA30}$ reduced SIDS risk 0.945 times, while a 1 knot increase in $South_{meanday-7}$ decreased SIDS risk by 0.927 times. This model implied that the greatest risk of SIDS occurred in winter, seven days after a day with little southerly winds.

Within day effects: The single within day effect found to be related to the risk of SIDS was $South_{rangeday-7}$. With a deviance statistic of $G = 5.583$, the addition of $South_{rangeday-7}$ significantly improved the baseline model ($p = 0.018$). An increase in the range of southern wind velocity a week ago resulted in a decreased risk of SIDS.

Between day effects: Significant associations with SIDS were found with three between day effects, namely $South_{maxday0-5}$, $South_{stdday0-5}$ and $South_{rangeday0-5}$. A preference for $South_{stdday0-5}$ over $South_{rangeday0-5}$ to be used in further analysis, is based on a comparison of goodness-of-fit statistics, with $South_{stdday0-5}$ yielding lower values on two of the three measures. An increase in both $South_{maxday0-5}$ and $South_{stdday0-5}$ related to an increase in SIDS risk.

Average day effects: No significant associations were found between any of the southern component of wind velocity and SIDS rates.

Between average day effects: As with the average day effects, no significant relationships were identified between SIDS and southerly wind velocity.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.056 (0.030)	0.945 (0.891, 1.003)			
$East_{maxAHCday-3-4}$	-0.097 (0.051)	0.907 (0.821, 1.002)	628.037	1800.318	6.707
$Temp_{meanMA30}$	-0.055 (0.030)	0.947 (0.892, 1.005)			
$East^{*}_{meanAHCday0-6}$	0.439 (0.176)	1.551 (1.099, 2.188)	626.322	1797.362	8.331
$Temp_{meanMA30}$	-0.059 (0.030)	0.943 (0.889, 1.000)			
$East^{*}_{stdAHCday-3-4}$	-0.434 (0.201)	0.648 (0.437, 0.962)	626.850	1806.530	5.982

Table C.7: Results from logistic regression of Wind Velocity (East), Period 1 ('*' = retained for further analysis).

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.066 (0.031)	0.936 (0.882, 0.994)			
$West^{*}_{meanday-7}$	-0.101 (0.051)	0.904 (0.819, 0.999)	627.646	1787.204	15.020
$Temp_{meanMA30}$	-0.061 (0.030)	0.941 (0.887, 0.998)			
$West^{*}_{maxday-2}$	-0.037 (0.019)	0.964 (0.928, 1.001)	628.371	1830.082	11.775
$Temp_{meanMA30}$	-0.061 (0.030)	0.941 (0.887, 0.998)			
$West^{*}_{rangeday-2}$	-0.038 (0.020)	0.963 (0.926, 1.001)	628.270	1830.186	9.094
$Temp_{meanMA30}$	-0.054 (0.030)	0.948 (0.893, 1.005)			
$West^{*}_{maxday0-1}$	-0.051 (0.027)	0.951 (0.902, 1.002)	628.290	1792.899	8.154
$Temp_{meanMA30}$	-0.053 (0.030)	0.949 (0.894, 1.007)			
$West^{*}_{stdday0-1}$	-0.173 (0.086)	0.841 (0.711, 0.994)	627.620	1809.897	6.059
$Temp_{meanMA30}$	-0.060 (0.030)	0.942 (0.887, 1.000)			
$West^{*}_{stdday0-3}$	0.125 (0.063)	1.133 (1.001, 1.283)	628.468	1805.454	9.462
$Temp_{meanMA30}$	-0.051 (0.030)	0.950 (0.896, 1.008)			
$West_{maxAHCday0-8}$	-0.064 (0.033)	0.938 (0.880, 1.000)	627.802	1795.238	5.510
$Temp_{meanMA30}$	-0.058 (0.030)	0.944 (0.890, 1.002)			
$West_{meanAHCday0-1}$	-0.318 (0.165)	0.728 (0.527, 1.005)	628.046	1801.201	8.394
$Temp_{meanMA30}$	-0.057 (0.031)	0.945 (0.890, 1.003)			
$West^{*}_{meanAHCday-4-5}$	0.301 (0.124)	1.351 (1.059, 1.724)	626.776	1802.402	12.496
$Temp_{meanMA30}$	-0.050 (0.030)	0.951 (0.897, 1.009)			
$West^{*}_{stdAHCday0-8}$	-0.285 (0.129)	0.752 (0.584, 0.969)	626.693	1793.566	4.773
$Temp_{meanMA30}$	-0.067 (0.030)	0.935 (0.881, 0.992)			
$West^{*}_{meanday-1to-2}$	0.127 (0.061)	0.881 (0.781, 0.993)	627.314	1809.717	4.321
$Temp_{meanMA30}$	-0.069 (0.031)	0.934 (0.879, 0.991)			
$West^{*}_{meanday-1to-3}$	-0.141 (0.069)	0.869 (0.759, 0.994)	627.512	1795.738	5.054

Table C.8: Results from logistic regression of Wind Velocity (West), Period 1 ('*' = retained for further analysis).

C.1.6 Wind Velocity — East

Of all the daily variables measuring the easterly component of wind velocity, only three were significantly associated with SIDS incidence. These were all between day effects: $East_{maxAHCday-3-4}$, $East_{meanAHCday0-6}$ and $East_{stdAHCday-3-4}$. The resulting models are presented in Table C.7. Of the two effects measuring the difference in within day effects from three to four days ago, $East_{stdAHCday-3-4}$, with lower D and HL statistics than $East_{maxAHCday-3-4}$, is retained for further analysis. The results from these models implied an increased SIDS risk with increased $East_{meanAHCday0-6}$, while conversely, an increase in $East_{stdAHCday-3-4}$ corresponded to a decreased risk of SIDS.

C.1.7 Wind Velocity — West

Models containing the western component of wind velocity that are significantly related to SIDS rates are presented in Table C.8. Of these variables, none had a significant seasonal distribution, that is, they were not related to $Temp_{meanMA30}$.

Daily summary measures: Two daily summary measures of westerly wind velocity were found to be associated with SIDS, over and above the baseline model: $West_{meanday-7}$ and $West_{maxday-2}$. With deviance statistics of $G = 6.21$ ($p = 0.013$) and $G = 5.485$ ($p = 0.019$) respectively, the addition of either variable to the seasonality model significantly improved the log-likelihood. Both variables corresponded to a decreased SIDS risk with increasing lagged westerly wind velocity.

Within day effects: The range of westerly wind velocity two days before *day*0 ($West_{rangeday-2}$) was the only within day effect to be significantly associated with SIDS incidence. Interpretation of this model

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.176 (0.106)	0.839 (0.682, 1.032)			
$WindC^{d}_{minday0}$	0.309 (0.101)	1.361 (1.116, 1.660)	189.261	557.104	5.698
$Temp_{meanMA30}$	-0.180 (0.106)	0.836 (0.679, 1.027)			
$WindC^{d*}_{meanday0}$	0.502 (0.154)	1.651 (1.222, 2.232)	187.635	528.175	4.254
$Temp_{meanMA30}$	-0.199 (0.104)	0.819 (0.668, 1.005)			
$WindC_{maxday0}$	0.434 (0.209)	1.544 (1.024, 2.327)	194.316	538.750	11.282
$Temp_{meanMA30}$	-0.180 (0.104)	0.836 (0.681, 1.025)			
$WindC^{d*}_{rangeday0}$	-0.201 (0.105)	0.818 (0.666, 1.004)	195.429	571.264	11.906

Table C.9: Results from logistic regression of Wind Chill, Period 1 ('d' = deseasoned; '*' = retained for further analysis).

indicated the greatest risk of SIDS occurred in winter when there was little variation in westerly wind velocity two days before *day*0.

Between day effects: The models corresponding to the between day effects that were found to be significantly related to SIDS are detailed in Table C.8. In retaining variables for further analysis, $West_{stdday0-1}$ was preferred over $West_{meanAHCday0-1}$, and $West_{stdAHCday0-8}$ was preferred over $West_{maxAHCday0-8}$. In both cases, goodness-of-fit statistics were utilised in the decision making process. An increase in $West_{maxday0-1}$, $West_{stdday0-1}$ and $West_{stdAHCday0-8}$ corresponded to a decrease in SIDS risk, while an increased risk of SIDS was associated with an increase in $West_{stdday0-3}$ and $West_{meanAHCday-4-5}$.

Average day effects: The average westerly wind velocity over the two and three days prior to the day of interest ($West_{meanday-1to-2}$ and $West_{meanday-1to-3}$) were found to be significantly related to the incidence of SIDS. Interpretation of the model containing $West_{meanday-1to-2}$ indicated that SIDS risk decreased 0.935 times with every 1°C increase in $Temp_{meanMA30}$, while a 1 knot increase in $West_{meanday-1to-2}$ decreased SIDS risk 0.881 times. A similar result was found with the model containing $West_{meanday-1to-3}$.

Between average day effects: No significant associations were found between any of the between average day effects and SIDS incidence.

C.1.8 Wind Chill

Wind chill is described in detail in Section 1.8.6. It is a contrived variable with a functional dependence on temperature and wind speed, only defined for temperatures less than 5°C and wind speeds greater than 5 km per hour . Therefore, with many hours of undefined data, only daily summary and within day measures were calculated. Significant wind chill models are presented in Table C.9, where variables which have been seasonally adjusted are denoted with a superscript 'd'. Due to the large amount of undefined data, Deviance and Pearson goodness-of-fit measures are not comparable to those generated from models containing different climate measures.

Daily summary measures: The three daily summary measures calculated for wind chill on the day of interest (*day*0) were all significantly related to the rate of SIDS, over and above the baseline model. $WindC_{meanday0}$, with lower values across all the goodness-of-fit statistics as compared to $WindC_{minday0}$ and $WindC_{maxday0}$, is retained for further examination of the daily wind chill relationship with SIDS. The model indicated that (on days where wind chill is defined) the greatest risk of SIDS occurred in winter, on days with a high wind chill index.

Within day measures: $WindC_{rangeday0}$ was the only within day measure of wind chill that was significantly associated with SIDS incidence. The odds ratios for this model implied that (on days where wind chill is defined) a 1°C increase in $Temp_{meanMA30}$ decreases SIDS risk 0.836 times, while a 1°C increase in $WindC_{rangeday0}$ decreased SIDS risk 0.818 times.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.062 (0.030)	0.940 (0.886, 0.998)			
$Humid_{maxAHCday0-5}$	0.027 (0.013)	1.027 (1.002, 1.054)	628.190	1801.007	9.050
$Temp_{meanMA30}$	-0.065 (0.031)	0.937 (0.883, 0.995)			
$Humid^{*}_{stdAHCday-4-5}$	0.125 (0.066)	1.133 (0.996, 1.288)	628.872	1812.498	6.642
$Temp_{meanMA30}$	-0.063 (0.030)	0.939 (0.885, 0.996)			
$Humid^{*}_{stdAHCday0-5}$	0.136 (0.058)	1.146 (1.023, 1.283)	627.190	1798.688	12.460

Table C.10: Results from logistic regression of Relative Humidity, Period 1 ('*' = retained for further analysis).

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.059 (0.030)	0.943 (0.889, 1.000)			
$Pres^{*}_{stdday-3}$	-0.170 (0.090)	0.844 (0.708, 1.006)	628.225	1793.532	3.669
$Temp_{meanMA30}$	-0.057 (0.030)	0.944 (0.890, 1.001)			
$Pres^{*}_{maxAHCday-8}$	-0.756 (0.270)	0.469 (0.277, 0.796)	622.454	1773.919	5.326
$Temp_{meanMA30}$	-0.057 (0.030)	0.944 (0.890, 1.001)			
$Pres^{*}_{meanAHCday-7}$	-1.362 (0.722)	0.256 (0.062, 1.054)	628.175	1799.690	9.333
$Temp_{meanMA30}$	-0.059 (0.030)	0.943 (0.889, 1.000)			
$Pres_{stdAHCday-8}$	-2.950 (1.121)	0.052 (0.006, 0.471)	623.806	1783.582	3.345
$Temp_{meanMA30}$	-0.065 (0.030)	0.937 (0.883, 0.994)			
$Pres^{*}_{maxday-7-8}$	-0.078 (0.035)	0.925 (0.864, 0.989)	626.349	1771.191	5.628
$Temp_{meanMA30}$	-0.052 (0.030)	0.950 (0.895, 1.007)			
$Pres^{*}_{maxAHCday-7-8}$	-0.696 (0.294)	0.499 (0.281, 0.887)	625.318	1843.704	2.530
$Temp_{meanMA30}$	-0.055 (0.030)	0.947 (0.893, 1.004)			
$Pres^{*}_{meanAHCday-1-2}$	-1.790 (0.911)	0.167 (0.028, 0.996)	627.703	1794.051	9.496
$Temp_{meanMA30}$	-0.053 (0.030)	0.949 (0.895, 1.007)			
$Pres_{stdAHC-7-8}$	-2.576 (1.260)	0.076 (0.006, 0.899)	627.209	1844.380	5.506

Table C.11: Results from logistic regression of Pressure, Period 1 ('*' = retained for further analysis).

C.1.9 Relative Humidity

Of all the relative humidity variables examined on a daily time scale, only three variables significantly improved the log-likelihood of the seasonality only model. These were $Humid_{maxAHCday0-5}$ ($G = 5.666$, $p = 0.017$), $Humid_{stdAHCday0-5}$ ($G = 6.666$, $p = 0.010$) and $Humid_{stdAHCday-4-5}$ ($G = 4.984$, $p = 0.026$), all between day measures of daily variation in humidity. The resulting models are summarised in Table C.10. Of the two variables measuring the difference in the variability in daily humidity between the day of interest and five days prior, $Humid_{stdAHCday0-5}$, with lower D and P statistics than $Humid_{maxAHCday0-5}$, is retained for further examination of the relationship between SIDS and humidity. Interpretation of these models showed that the risk of SIDS increased as $Humid_{stdAHCday-4-5}$ or $Humid_{stdAHCday0-5}$ increased.

C.1.10 Pressure

The results for the models corresponding to a significant association between pressure variables and SIDS incidence are presented in Table C.11. There was no apparent seasonal trend in the pressure variables.

Daily summary measures: No significant relationships were identified between any of the daily summary measures of pressure and the rate of SIDS.

Within day effects: Four variables measuring the variation in daily pressure were found to be associated with the incidence of SIDS, namely $Pres_{stdday-3}$, $Pres_{maxAHCday-8}$, $Pres_{meanAHCday-7}$ and $Pres_{stdAHCday-8}$. The addition of each of these variables significantly improved the log-likelihood of the baseline seasonality model (G ranging from 5.631 ($p = 0.018$) to 11.402 ($p = 0.001$)). Of the two variables measuring the daily pressure variations eight days before $day0$, $Pres_{maxAHCday-8}$ is the preferred variable due to its resulting goodness-of-fit statistics, and is therefore retained for further analysis. All of the significant within day pressure effects showed that an increase in the variability of pressure over the past few days resulted in a decreased SIDS risk.

Between day effects: Four between day measures of pressure were found to be significantly related to SIDS incidence, over and above seasonality. The resulting models are presented in Table C.11.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.058 (0.030)	0.943 (0.889, 1.001)			
$Rain^{*}_{maxday-1}$	-0.368 (0.194)	0.692 (0.473, 1.013)	626.850	1772.501	8.307
$Temp_{meanMA30}$	-0.059 (0.030)	0.943 (0.889, 1.000)			
$Rain^{*}_{stdday-1}$	-1.558 (0.779)	0.211 (0.046, 0.969)	628.048	1768.241	3.463
$Temp_{meanMA30}$	-0.058 (0.030)	0.943 (0.889, 1.001)			
$Rain_{rangeday-1}$	-0.368 (0.194)	0.692 (0.473, 1.013)	626.852	1772.540	6.269
$Temp_{meanMA30}$	-0.058 (0.030)	0.943 (0.889, 1.000)			
$Rain_{meanAHCday-1}$	-3.467 (1.770)	0.031 (< 0.001, 1.001)	626.134	1782.608	6.384
$Temp_{meanMA30}$	-0.056 (0.030)	0.945 (0.891, 1.034)			
$Rain^{*}_{stdday0-1}$	-1.134 (0.578)	0.322 (0.104, 0.999)	626.963	1763.281	9.704
$Temp_{meanMA30}$	-0.055 (0.030)	0.946 (0.892, 1.004)			
$Rain_{maxAHCday0-1}$	-0.370 (0.189)	0.691 (0.477, 1.001)	626.826	1777.892	10.264
$Temp_{meanMA30}$	-0.055 (0.030)	0.946 (0.892, 1.004)			
$Rain_{stdAHCday0-1}$	-1.315 (0.690)	0.268 (0.069, 1.038)	627.192	1779.617	6.863

Table C.12: Results from logistic regression of Rainfall, Period 1 ('*' = retained for further analysis).

$Pres_{maxAHCday-7-8}$ is retained over $Pres_{stdAHCday-7-8}$ as it has lower values across all goodness-of-fit measures. Again, an increase in all variables was associated with a decreased risk of SIDS; the odds ratio for the $Pres_{maxAHCday-7-8}$ model implied that the risk of SIDS halved with a 1 HPa increase in pressure.

Average day effects: No significant associations between the average pressure over the few days prior to the day of interest and SIDS incidence were found.

Between average day effects: Like the average day effects, no significant relationships were found between the rate of SIDS and the difference in pressure on the day of interest and the past average, over and above the seasonal component of the SIDS distribution.

C.1.11 Rainfall

Rainfall was not significantly associated with $Temp_{meanMA30}$, confirming the lack of seasonality visually observed in Figure 1.17 in Chapter 1.8.

Daily summary measures: A single daily summary measure of rainfall was found to be related to the incidence of SIDS, namely $Rain_{maxday-1}$, the maximum rainfall in one hour on the day prior to the day of interest. With a deviance statistic of $G = 7.006$ ($p = 0.008$), there is sufficient evidence to show that the addition of $Rain_{maxday-1}$ to the baseline seasonality model significantly improves the log-likelihood. Interpretation of the corresponding odds ratios for the model implied that SIDS risk decreased 0.943 times with each 1°C increase in $Temp_{meanMA30}$, while a 1mm increase in $Rain_{maxday-1}$ resulted in the risk of SIDS decreasing 0.692 times.

Within day effects: Three within day rainfall effects were significantly associated with SIDS rate. $Rain_{stdday-1}$ was preferred over $Rain_{rangeday-1}$ and $Rain_{meanAHCday-1}$ as a measure of the variation in rainfall, as it had lower values for both the P and HL goodness-of-fit statistics. The model shows that the greatest risk of SIDS occurred in winter where there was little variation in the rainfall on the day before $day0$.

Between day effects: Again, three measures of the difference in rainfall between days were found to be significantly associated with SIDS, namely $Rain_{stdday0-1}$, $Rain_{maxAHCday0-1}$ and $Rain_{stdAHCday0-1}$. $Rain_{stdday0-1}$, with lower values on two of the three goodness-of-fit statistics, as compared to $Rain_{maxAHCday0-1}$ and $Rain_{stdAHCday0-1}$, is retained for further analysis. An increase in $Rain_{stdday0-1}$ implied a decrease in SIDS risk.

Average day effects: No significant relationships between average rainfall over the past days and the incidence of SIDS were found.

Between average day effects: Similarly, there was no evidence of any between average rainfall day effects having any impact on the rate of SIDS occurrence.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	-0.069 (0.031)	0.933 (0.878, 0.992)			
$Sun^{*}_{totalday-2-3}$	0.071 (0.037)	1.074 (0.999, 1.155)	628.596	1789.834	4.627
$Temp_{meanMA30}$	-0.065 (0.031)	0.937 (0.883, 0.995)			
$Sun^{*}_{totalday0-7}$	0.079 (0.035)	1.083 (1.010, 1.160)	627.237	1808.267	5.114

Table C.13: Results from logistic regression of Sunshine, Period 1 ('*' = retained for further analysis).

C.1.12 Sunshine

As noted in Section 1.8.10, the sunshine data was only available as the total hours of sunshine per day. This restricted the number of daily measures of sunshine that were able to be created. Sun variables examined cover the total number of sunshine hours per day for the day of interest and lags, the difference in the total number of sunshine hours per day on consecutive days, and from $day0$ and previous lags, the average total sunshine per day over the past few days, and the difference between this and the total number of sunshine hours on the day of interest ($day0$).

Of those variables created, only two were found to be significantly associated with the incidence of SIDS, over and above seasonality: $Sun_{totalday-2-3}$ and $Sun_{totalday0-7}$. Both variables were between day effects and implied that SIDS risk increased with increasing $Sun_{totalday-2-3}$ or $Sun_{totalday0-7}$. The $Sun_{totalday0-7}$ model inferred that the greatest risk of SIDS occurs in winter when the number of hours of sunshine on $day0$ was greater than it was a week prior.

C.1.13 Solar Radiation

As with sunshine, the variables created to measure solar radiation were restricted to those based on the total global solar radiation per day. Of these variables, only the between average day effects, the difference between the total radiation on $day0$ and past averages, were found to not have a significant seasonal component in their underlying distribution. After adjustment for seasonality, no relationship between solar radiation and SIDS incidence was identified.

C.1.14 Dewpoint

Dewpoint showed no significant associations with $Temp_{meanMA30}$, that is, there was no apparent underlying seasonal pattern in the dewpoint variables. However, as with wind speed, the addition of most daily dewpoint variables to the baseline seasonality model resulted in $Temp_{meanMA30}$ no longer being significant. Table C.14 presents the models corresponding to variables found to be significantly related to the incidence of SIDS.

Daily summary measures: The single daily summary of dewpoint found to be related to SIDS rates was $Dew_{minday-4}$. With a corresponding deviance statistic $G = 4.079$ ($p = 0.043$), the addition of $Dew_{minday-4}$ to the seasonality only model significantly improved the log-likelihood. This model inferred an decrease in SIDS risk with increased minimum dewpoint temperature four days prior.

Within day effects: No significant associations between the incidence of SIDS and any variation in daily dewpoint measures were found.

Between day effects: Most of the significant associations between measures of dewpoint and SIDS involved between day effects. The complete list of models containing significant dewpoint between day covariates is presented in Table C.14. Those that are retained for further analysis are: $Dew_{meanday-1-2}$, $Dew_{meanday0-2}$, $Dew_{meanday0-14}$, $Dew_{maxday0-3}$, $Dew_{stdday-2-3}$, $Dew_{rangeday0-2}$ and $Dew_{maxAHCday0-5}$. Goodness-of-fit statistics were used when deciding between competing models with covariates essentially

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$Temp_{meanMA30}$	0.048 (0.031)	0.953 (0.896, 1.013)			
$Dew_{minday-4}$	-0.051 (0.025)	1.052 (1.001, 1.105)	629.777	1801.939	6.949
$Temp_{meanMA30}$	0.043 (0.031)	0.958 (0.901, 1.018)			
$Dew_{minday-1-2}$	-0.134 (0.067)	1.144 (1.004, 1.304)	629.305	1803.630	7.797
$Temp_{meanMA30}$	0.041 (0.031)	0.960 (0.904, 1.019)			
$Dew_{minday0-2}$	-0.129 (0.055)	1.137 (1.022, 1.266)	627.637	1796.518	7.015
$Temp_{meanMA30}$	0.043 (0.031)	0.958 (0.901, 1.017)			
$Dew^{*}_{meanday-1-2}$	-0.205 (0.100)	1.227 (1.009, 1.492)	629.146	1787.489	20.038
$Temp_{meanMA30}$	0.043 (0.031)	0.958 (0.902, 1.018)			
$Dew^{*}_{meanday0-2}$	-0.279 (0.083)	1.322 (1.125, 1.554)	620.049	1766.019	5.456
$Temp_{meanMA30}$	0.042 (0.031)	0.959 (0.903, 1.019)			
$Dew_{meanday0-3}$	-0.139 (0.066)	1.149 (1.010, 1.306)	628.936	1782.182	10.537
$Temp_{meanMA30}$	0.043 (0.031)	0.958 (0.902, 1.018)			
$Dew^{*}_{meanday0-14}$	-0.139 (0.060)	1.149 (1.022, 1.292)	627.814	1801.422	8.532
$Temp_{meanMA30}$	0.042 (0.031)	0.959 (0.903, 1.019)			
$Dew_{maxday0-2}$	-0.206 (0.084)	1.228 (1.042, 1.448)	627.061	1763.864	8.629
$Temp_{meanMA30}$	0.042 (0.031)	0.959 (0.903, 1.019)			
$Dew^{*}_{maxday0-3}$	-0.219 (0.079)	1.244 (1.066, 1.452)	625.088	1761.643	7.195
$Temp_{meanMA30}$	0.044 (0.031)	0.957 (0.901, 1.017)			
$Dew_{maxday0-14}$	-0.146 (0.068)	1.157 (1.013, 1.322)	628.719	1803.897	7.197
$Temp_{meanMA30}$	0.043 (0.031)	0.958 (0.902, 1.018)			
$Dew^{*}_{stdday-2-3}$	-0.583 (0.233)	1.791 (1.134, 2.829)	626.541	1783.101	3.619
$Temp_{meanMA30}$	0.044 (0.031)	0.957 (0.901, 1.017)			
$Dew_{rangeday-2-3}$	-0.141 (0.069)	1.151 (1.005, 1.319)	629.229	1797.765	7.519
$Temp_{meanMA30}$	0.043 (0.031)	0.959 (0.901, 1.017)			
$Dew^{*}_{rangeday0-2}$	-0.131 (0.063)	1.140 (1.007, 1.291)	629.065	1810.125	7.719
$Temp_{meanMA30}$	0.044 (0.031)	0.957 (0.901, 1.017)			
$Dew^{*}_{maxAHCday0-5}$	0.147 (0.061)	0.864 (0.767, 0.973)	628.746	1793.573	12.695
$Temp_{meanMA30}$	0.043 (0.031)	0.958 (0.901, 1.018)			
$Dew_{stdAHCday0-5}$	0.620 (0.292)	0.538 (0.304, 0.952)	629.820	1794.904	6.085
$Temp_{meanMA30}$	0.040 (0.031)	0.960 (0.904, 1.020)			
$Dew^{*}_{diffday0-14}$	0.087 (0.045)	0.917 (0.840, 1.000)	629.931	1774.763	11.123

Table C.14: Results from logistic regression of Dewpoint, Period 1 ('*' = retained for further analysis).

capturing the same climatic profiles. With the exception of $Dew_{maxAHCday0-5}$, an increase in each of the between day effects corresponded to a decrease in the risk of SIDS.

Average day effects: No average day effects of dewpoint were identified as being significantly associated with the incidence of SIDS.

Between average day effects: The difference between the mean dewpoint on *day*0 and the average dewpoint over the fourteen days before *day*0 ($Dew_{diffday0-14}$) was the only between average day effect to be significantly related to SIDS. This model implied that the greatest risk of SIDS occurred in winter on days were the dewpoint temperature was more than it had been over the past fourteen days.

C.1.15 Multiple Logistic Regression Models

Table C.15 shows the results from a multiple logistic model. All 125 candidate climate variables were initially included in the model, and a backwards elimination procedure used to give the resulting model. Significance was taken as $\alpha \leqslant 0.05$. This model shows a multi-factorial relationship between SIDS and climate; the model incorporates variants from the complete suite of available climate variables. Principal component regression (PCR, see Chapter 4 for details) will be used to further examine the SIDS-climate daily relationship in Period 1. PCR will enable the complete climatic structure to be examined in relation to SIDS, and potentially highlight a more detailed, informative climate SIDS model. The advantage of PCR is that it inherently allows for the interrelationships within the predictor variables.

Variable	Estimate (SE)	Odds ratio (95% CI)
Intercept	2.93 (-0.72)	
$temp_{meanMA30}$	0.07 (0.03)	1.07 (1.00, 1.15)
$East_{stdAHCday-3-4}$	0.54 (0.22)	1.72 (1.11, 2.65)
$North_{maxday0}$	-0.08 (0.02)	0.92 (0.88, 0.97)
$North_{minday-7}$	-0.35 (0.15)	0.71 (0.53, 0.96)
$South_{maxAHCday0-4}$	0.07 (0.03)	1.07 (1.01, 1.15)
$South_{maxday0-5}$	-0.04 (0.02)	0.96 (0.92, 1.00)
$West_{meanAHCday-4-5}$	-0.38 (0.14)	0.69 (0.52, 0.90)
$West_{stdAHCday0-8}$	0.35 (0.14)	1.41 (1.07, 1.88)
$West_{stdday0-1}$	0.25 (0.10)	1.28 (1.06, 1.55)
$West_{stdday0-3}$	-0.16 (0.08)	0.86 (0.74, 0.99)
$Dew_{maxday0-3}$	-0.24 (0.07)	0.78 (0.68, 0.90)
$Dew_{meanday0-2}$	0.16 (0.08)	1.17 (1.01, 1.36)
$Humid_{stdAHC-4-5}$	-0.16 (0.07)	0.85 (0.74, 0.98)
$Pres_{maxAHCday-7-8}$	0.94 (0.32)	2.56 (1.36, 4.85)
$Pres_{meanAHCday-1-2}$	2.75 (1.04)	15.57 (2.04, 119.06)
$Pres_{meanday-5-6}$	0.07 (0.04)	1.08 (1.00, 1.16)
$Pres_{rangeday0}$	-0.06 (0.03)	0.94 (0.89, 0.99)
$Pres_{rangeday-14}$	0.07 (0.03)	1.07 (1.01 1.14)
$Rain_{meanAHCday0-5}$	2.49 (1.24)	12.04 (1.07, 135.66)
$Rain_{stdday-14}$	-0.81 (0.33)	0.44 (0.23, 0.86)
$Sun_{totalday-2-3}$	-0.09 (0.04)	0.92 (0.84, 0.99)
$Temp_{meanday0-2}$	0.15 (0.07)	1.16 (1.01, 1.34)
$Temp_{stdday-7}$	-0.26 (0.09)	0.77 (0.65, 0.92)
$Temp_{stdday0-5}$	-0.25 (0.11)	0.78 (0.63, 0.96)
$WindD_{maxAHCday-4}$	0.01 (0.00)	1.01 (1.00, 1.02)
$WindS_{meanday-1-2}$	-0.09 (0.04)	0.92 (0.84, 1.00)

Table C.15: Results from multiple logistic regression of daily climate variables, Period 1.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.504 (0.071)	0.604 (0.525, 0.694)			
$Temp_{maxAHCday-8}$	0.058 (0.028)	1.060 (1.002, 1.120)	3246.267	6185.727	5.542
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.796 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.505 (0.071)	0.604 (0.525, 0.694)			
$Temp^{*}_{stdAHCday-8}$	0.312 (0.132)	1.366 (1.054, 1.771)	3245.006	6185.083	5.536
$sin(\frac{2\pi t}{365})$	-0.231 (0.070)	0.794 (0.693, 0.910)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.071)	0.593 (0.516, 0.682)			
$Temp^{*}_{meanday0-14}$	0.044 (0.019)	1.045 (1.006, 1.085)	3245.299	6191.100	2.408
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.696, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Temp^{*}_{rangeday0-5}$	0.028 (0.014)	1.028 (1.000, 1.057)	3246.565	6189.276	4.876
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.508 (0.071)	0.601 (0.523, 0.691)			
$Temp^{*}_{maxAHCday0-8}$	0.068 (0.030)	1.071 (1.009, 1.136)	3245.502	6175.475	11.440

Table C.16: Results from logistic regression of temperature, Period 2 ('*' = retained for further analysis).

C.2 Period 2 (1973—1989)

C.2.1 Temperature

Logistic regression models containing daily temperature measures that were significantly related to the incidence of SIDS in Period 2, over and above seasonality, are presented in Table C.16. Variables that are retained for further analysis are superscripted with a '*'. No significant relationships were found between SIDS and any of the temperature daily summary measures, average day effects or between average day effects.

Within day effects: Two measures of the variability in the temperature eight days prior to the day of interest were found to be related to SIDS, namely $Temp_{maxAHCday-8}$ and $Temp_{stdAHCday-8}$. Of these two variables, $Temp_{stdAHCday-8}$, with lower D, P, and HL statistic values, is retained for further analysis. This model implied that SIDS risk increased 1.366 times with every 1°C increase in $Temp_{stdAHdayC-8}$; that is, the greatest risk of SIDS occurred in winter, eight days after a day where the temperature was highly variable.

Average Day effects	Estimate (SE) $WindD_{day-3}$	Estimate (SE) $WindD_{day-4}$	Estimate (SE) $WindD_{day-7}$
$sin(\frac{2\pi t}{365})$	-0.231 (0.070)	-0.227 (0.070)	-0.225 (0.070)
$cos(\frac{2\pi t}{365})$	-0.481 (0.079)	-0.545 (0.079)	-0.460 (0.079)
1	0.105 (0.191)	0.261 (0.216)	0.497* (0.209)
2	0.058 (0.292)	0.416 (0.306)	-0.146 (0.357)
3	0.005 (0.199)	0.451* (0.217)	0.197 (0.221)
4	0.021 (0.189)	0.299 (0.213)	0.152 (0.213)
5	-8.734 (335.9)	-8.351 (335.9)	-8.450 (335.9)
6	-0.572* (0.282)	0.459 (0.254)	0.231 (0.261)
7	-0.215 (0.202)	0.161 (0.221)	0.295 (0.214)
8	-0.150 (0.243)	0.185 (0.260)	0.146 (0.260)
9	-0.437 (0.282)	0.463 (0.261)	0.515* (0.254)
11 (*reference category*)	0	0	0
Deviance	3236.766	3242.506	3238.118
Pearson	6170.338	6197.627	6189.622
Hosmer-Lemeshow	8.816	6.863	11.712

Table C.17: Results from logistic regression of Wind Direction, Period 2.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.219 (0.070)	0.804 (0.701, 0.921)			
$cos(\frac{2\pi t}{365})$	-0.495 (0.072)	0.610 (0.530, 0.701)			
$WindD^*_{maxAHCday-4-5}$	-0.003 (0.001)	0.997 (0.994, 1.000)	3244.924	6176.463	11.274
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.509 (0.071)	0.601 (0.523, 0.691)			
$WindD^*_{stdAHCday0-3}$	-0.010 (0.005)	0.990 (0.980, 1.000)	3246.322	6178.050	6.518

Table C.18: Results from logistic regression of Wind Direction, Period 2 ('*' = retained for further analysis).

Between day effects: Three measures of between day temperature effects were found to be associated with SIDS: $Temp_{meanday0-14}$, $Temp_{rangeday0-5}$ and $Temp_{maxAHCday0-8}$. With deviance statistics ranging from $G = 5.059$ ($p = 0.025$) to $G = 3.793$ ($p = 0.051$) for these three models, inclusion of any of the variables significantly improved the log-likelihood of the baseline seasonality only model. An increase in any of these temperature measures implied an increased risk of SIDS.

C.2.2 Wind Direction

Results from the analysis of wind direction in relation to the incidence of SIDS appear in Table C.17 and Table C.18. Three lagged variables relating to the predominant wind direction were found to be significantly associated with SIDS, over and above seasonality, as shown in Table C.17. Taking the 'mixed' category (no one wind direction predominating) as the reference category, SIDS risk decreased 0.564 times when the predominant wind direction three days prior was north-east. In contrast, an easterly wind four days before *day*0, or a calm day or north-westerly wind one week before *day*0, were all related to an increased risk of SIDS.

The significant relationships between the variables measuring absolute hourly change in wind direction and SIDS are presented in Table C.18. Only two variables were significantly related to SIDS incidence, over and above seasonality, namely $WindD_{maxAHCday-4-5}$ and $WindD_{stdAHCday0-3}$. An increase in either variable corresponded to a decrease in the risk of SIDS.

C.2.3 Wind Speed

Table C.19 presents results from the analysis of SIDS incidence in relation to daily wind speed measures with only models containing significant covariates presented. There were no significant associations between wind speed and seasonality as measured by the first order sinusoid functional.

Daily summary measures: One measure of daily wind speed was found to be related to SIDS over and above the seasonal fluctuations, namely $WindS_{maxday0}$, the maximum wind speed on the day of interest.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.802 (0.700, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.511 (0.071)	0.600 (0.522, 0.690)			
$WindS^{*}_{maxday0}$	-0.025 (0.009)	0.975 (0.958, 0.992)	3241.737	6196.996	8.363
$sin(\frac{2\pi t}{365})$	-0.216 (0.070)	0.806 (0.703, 0.923)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.601 (0.522, 0.691)			
$WindS^{*}_{stdday0}$	-0.037 (0.010)	0.964 (0.946, 0.982)	3238.455	6203.447	2.793
$sin(\frac{2\pi t}{365})$	-0.216 (0.070)	0.806 (0.703, 0.923)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.601 (0.522, 0.691)			
$WindS_{rangeday0}$	-0.037 (0.010)	0.964 (0.946, 0.982)	3235.204	6217.244	5.334
$sin(\frac{2\pi t}{365})$	-0.220 (0.070)	0.802 (0.700, 0.920)			
$cos(\frac{2\pi t}{365})$	-0.513 (0.071)	0.599 (0.521, 0.688)			
$WindS_{stdAHCday0}$	-0.126 (0.064)	0.882 (0.778, 0.001)	3246.273	6215.932	6.268
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$WindS_{minday-7-8}$	0.038 (0.017)	1.039 (1.005, 1.073)	3245.425	6187.125	11.106
$sin(\frac{2\pi t}{365})$	-0.217 (0.070)	0.805 (0.702, 0.923)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.600 (0.521, 0.689)			
$WindS^{*}_{minday0-8}$	0.039 (0.015)	1.039 (1.009, 1.071)	3244.284	6196.863	9.814
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.802 (0.699, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.598 (0.520, 0.687)			
$WindS^{*}_{meanday-7-8}$	0.043 (0.018)	1.044 (1.009, 1.081)	3244.569	6191.277	6.814
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.802 (0.699, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$WindS^{*}_{stdday0-5}$	-0.108 (0.039)	0.897 (0.831, 0.969)	3242.284	6215.968	17.101
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$WindS_{stdday0-8}$	-0.073 (0.038)	0.929 (0.863, 1.001)	3246.407	6194.082	9.059
$sin(\frac{2\pi t}{365})$	-0.217 (0.070)	0.805 (0.702, 0.922)			
$cos(\frac{2\pi t}{365})$	-0.513 (0.071)	0.599 (0.521, 0.688)			
$WindS_{maxAHCday-5-6}$	-0.036 (0.017)	0.964 (0.933, 0.996)	3245.252	6214.401	6.176
$sin(\frac{2\pi t}{365})$	-0.218 (0.070)	0.804 (0.701, 0.921)			
$cos(\frac{2\pi t}{365})$	-0.511 (0.071)	0.600 (0.522, 0.690)			
$WindS^{*}_{maxAHCday0-3}$	-0.030 (0.016)	0.971 (0.942, 1.001)	3246.460	6219.303	12.045
$sin(\frac{2\pi t}{365})$	-0.216 (0.070)	0.806 (0.703, 0.924)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.600 (0.522, 0.690)			
$WindS^{*}_{maxAHCday0-5}$	-0.037 (0.016)	0.964 (0.935, 0.994)	3244.340	6246.747	1.375
$sin(\frac{2\pi t}{365})$	-0.219 (0.070)	0.803 (0.701, 0.921)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$WindS^{*}_{maxAHCday0-6}$	-0.033 (0.016)	0.968 (0.939, 0.998)	3245.584	6226.943	8.929
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$WindS^{*}_{meanAHCday-5-6}$	-0.173 (0.079)	0.841 (0.720, 0.982)	3245.254	6205.778	4.952
$sin(\frac{2\pi t}{365})$	-0.219 (0.070)	0.804 (0.701, 0.921)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.600 (0.522, 0.690)			
$WindS^{*}_{meanAHCday0-8}$	-0.167 (0.066)	0.846 (0.743, 0.963)	3243.600	6227.630	6.909
$sin(\frac{2\pi t}{365})$	-0.217 (0.070)	0.805 (0.702, 0.923)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.598 (0.520, 0.687)			
$WindS_{stdAHCday-5-6}$	-0.226 (0.081)	0.798 (0.682, 0.934)	3241.810	6216.512	8.130
$sin(\frac{2\pi t}{365})$	-0.215 (0.070)	0.807 (0.704, 0.925)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.600 (0.522, 0.690)			
$WindS_{stdAHCday0-5}$	-0.205 (0.073)	0.814 (0.706, 0.939)	3241.737	6261.773	4.344

Table C.19: Results from logistic regression of Wind Speed, Period 2 ('*' = retained for further analysis).

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.071)	0.593 (0.515, 0.682)			
$North^{*}_{minday-14}$	0.180 (0.074)	1.197 (1.035, 1.385)	3245.348	6195.992	11.431
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.478 (0.073)	0.620 (0.538, 0.715)			
$North^{*}_{maxday-7}$	-0.021 (0.010)	0.980 (0.961, 0.998)	3245.683	6198.841	2.300
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.480 (0.073)	0.619 (0.537, 0.714)			
$North^{*}_{rangeday-7}$	-0.020 (0.010)	0.980 (0.961, 0.999)	3246.078	6199.438	3.026
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.801 (0.699, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.483 (0.073)	0.617 (0.535, 0.711)			
$North_{maxAHCday-7}$	-0.028 (0.014)	0.972 (0.946, 1.000)	3246.239	6203.191	2.045
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.802 (0.700, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.481 (0.073)	0.618 (0.536, 0.713)			
$North_{stdAHCday-7}$	-0.115 (0.053)	0.892 (0.803, 0.990)	3245.577	6204.438	4.355
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.802 (0.700, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.524 (0.072)	0.592 (0.515, 0.681)			
$North_{minday0-14}$	0.124 (0.060)	1.132 (1.008, 1.272)	3246.445	6189.950	11.210
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.531 (0.072)	0.588 (0.511, 0.677)			
$North^{*}_{meanday0-14}$	0.052 (0.020)	1.054 (1.013, 1.096)	3243.835	6186.063	6.705
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.598 (0.520, 0.687)			
$North^{*}_{stdday0-8}$	0.063 (0.033)	1.065 (0.999, 1.136)	3246.689	6175.727	12.772

Table C.20: Results from logistic regression of Wind Velocity (North), Period 2 ('*' = retained for further analysis).

With a corresponding deviance statistic of $G = 8.621$ ($p = 0.003$), the addition of $WindS_{maxday0}$ significantly improved the log-likelihood over the baseline only model. This model implied that a 1 knot increase in the maximum wind speed corresponded to a drop in the risk of SIDS by 0.975 times.

Within day effects: Three models containing variables measuring the variation in wind speed on the day of interest were identified as significantly associated with the rate of SIDS. $WindS_{stdday0}$ is retained, over $WindS_{rangeday0}$ or $WindS_{stdAHCday0}$, as it had lower values across all three goodness-of-fit statistics. This model implied that the greatest risk of SIDS occurred in winter, on days where the wind was reasonably constant.

Between day effects: Thirteen within day effects were found to be significantly related to the incidence of SDIS, for both differences in consecutive days and differences between $day0$ and various lags. Full details are given in Table C.19. Goodness-of-fit statistics were utilised to decide which variables are retained for further analysis of the relationship between SIDS and wind speed. Successful candidates were $WindS_{meanday-7-8}$, $WindS_{minday0-8}$, $WindS_{meanAHCday0-8}$, $WindS_{stdday0-5}$, $WindS_{meanAHCday-5-6}$, $WindS_{maxAHCday0-3}$, $WindS_{maxAHCday0-5}$, and $WindS_{maxAHCday0-6}$. With the exception of $WindS_{meanday-7-8}$ and $WindS_{minday0-8}$, an increase in any of the retained variables corresponded to a decrease in the risk of SIDS.

Average day effects: No significant associations were found between any of the average day effects of wind speed and SIDS rates.

Between average day effects: As with the average day effects, no significant relationships were identified between SIDS and between average day effects of wind speed.

C.2.4 Wind Velocity - North

The northern component of wind velocity was not significantly associated with seasonality, implying that there is no underlying seasonal pattern in the variable. The results are presented in Table C.20 for those models corresponding to north wind velocity variables found to be significantly related to the rate of SIDS, over and above seasonality. No average or between average day effects were significant.

Daily summary measures: Two daily measures of the northerly wind velocity were significantly associated with SIDS incidence: the minimum north wind velocity two weeks before *day*0 ($North_{minday-14}$) and the maximum north wind velocity one week before *day*0 ($North_{maxday-7}$). An increase in $North_{minday-14}$ implied an increase in the risk of SIDS, while conversely, an increase in $North_{maxday-7}$ corresponded to a decreased SIDS risk.

Within day effects: Three measures of the variability in northerly wind velocity a week prior to the day of interest were found to be significantly related to SIDS, over and above seasonality, namely $North_{rangeday-7}$, $North_{maxAHCday-7}$ and $North_{stdAHCday-7}$. As all three variables are essentially capturing the same climatic profile, only $North_{rangeday-7}$, with lower scores on two of the three goodness-of-fit measures, is retained for further analysis. This model showed that a 1 knot increase in $North_{rangeday-7}$ decreased the risk of SIDS 0.980 times. This implied that the greatest risk of SIDS occurred in winter, when there was little variation in the northerly component of wind velocity one week prior.

Between day effects: The three between day effects corresponding to the northerly component of wind velocity that were found to be significantly associated with SIDS were $North_{minday0-14}$, $North_{meanday0-14}$ and $North_{stdday0-8}$. Of the two variables measuring the difference in north wind velocity between the day of interest and 14 days before the day of interest, $North_{meanday0-14}$, with lower values across all three goodness-of-fit statistics, is retained for further analysis. An increase in either $North_{meanday0-14}$ or $North_{stdday0-8}$ corresponded to an increase in the risk of SIDS.

C.2.5 Wind Velocity - South

As with wind speed and northerly wind velocity in Period 2, the southern component of wind velocity was not significantly associated with seasonality, as measured by a first order sinusoid. Table C.21 presents the models where a southerly wind velocity measure was significantly related to the incidence of SIDS over and above the seasonal component.

Daily summary measures: The mean southern wind velocity three and four days before *day*0, and the maximum southern wind velocity three days before *day*0 were all found to be significantly related to SIDS incidence. With lower values on two of the three goodness-of-fit statistics, $South_{maxday-3}$ is retained over $South_{meanday-3}$ for further examination of the relationship between southerly wind velocity three days before *day*0 and SIDS. Interpretation of the odds ratio for the model containing $South_{maxday-3}$ implied that a 1 knot increase in $South_{maxday-3}$ resulted in the risk of SIDS decreasing 0.980 times. Similarly for the model containing $South_{meanday-4}$, a 1 knot increase in $South_{meanday-4}$ resulted in a 0.965 decrease in SIDS risk.

Within day effects: All the variables measuring variation in southerly wind velocity three days before *day*0 were found to be significantly associated with the incidence of SIDS. Of the five variables, $South_{meanAHCday-3}$ is retained due to it having lower values of at least two of the three goodness-of-fit statistics, as compared to the other variables. This model suggested that the greatest risk of SIDS occurred in winter, on days where the wind velocity from the southerly direction varied widely three days prior to the day of interest.

Between day effects: Three between day effects were found to be related to the risk of SIDS, namely $South_{meanday-3-4}$, $South_{meanday-4-5}$ and $South_{maxday-7-8}$. With deviance statistics ranging from $G = 6.519$ ($p = 0.011$) to $G = 3.643$ ($p = 0.056$) for these models, the addition of any of the three variables significantly improved the baseline model. An increase in either $South_{meanday-3-4}$ or $South_{meanday-4-5}$ corresponded to a decrease in the risk of SIDS, whereas an increase in $South_{maxday-7-8}$ implied an increase in SIDS risk.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.796 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$South^{*}_{meanday-3}$	-0.054 (0.018)	0.948 (0.915, 0.981)	3240.427	6193.827	7.225
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.695, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$South^{*}_{meanday-4}$	-0.035 (0.017)	0.965 (0.934, 0.998)	3245.836	6189.805	7.708
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.508 (0.071)	0.602 (0.524, 0.692)			
$South^{*}_{maxday-3}$	-0.020 (0.007)	0.980 (0.966, 0.993)	3241.620	6189.897	6.804
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.503 (0.071)	0.605 (0.526, 0.695)			
$South^{*}_{stdday-3}$	-0.066 (0.023)	0.936 (0.895, 0.979)	3244.475	6193.979	8.971
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.507 (0.071)	0.603 (0.524, 0.693)			
$South^{*}_{rangeday-3}$	-0.020 (0.007)	0.980 (0..966 0.994)	3242.039	6190.071	7.235
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.801 (0.699, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.506 (0.071)	0.603 (0.524, 0.693)			
$South_{maxAHCday-3}$	-0.020 (0.010)	0.980 (0.961, 1.000)	3246.211	6182.102	5.729
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.518 (0.071)	0.595 (0.518, 0.685)			
$South^{*}_{meanAHCday-3}$	-0.152 (0.053)	0.859 (0.774, 0.954)	3241.898	6182.518	5.601
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.507 (0.071)	0.602 (0.524, 0.692)			
$South_{stdAHCday-3}$	-0.088 (0.041)	0.916 (0.845, 0.992)	3245.546	6182.434	8.324
$sin(\frac{2\pi t}{365})$	-0.225 (0.070)	0.799 (0.697, 0.915)			
$cos(\frac{2\pi t}{365})$	-0.497 (0.071)	0.608 (0.529, 0.700)			
$South_{meanday-3-4}$	-0.049 (0.020)	0.952 (0.915, 0.990)	3243.839	6188.471	4.188
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.499 (0.071)	0.607 (0.528, 0.698)			
$South^{*}_{meanday-4-5}$	-0.041 (0.020)	0.960 (0.923, 0.997)	3245.764	6179.7281	13.608
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.528 (0.072)	0.590 (0.512, 0.679)			
$South^{*}_{maxAHCday-7-8}$	0.019 (0.010)	1.019 (1.000, 1.038)	3246.715	6180.932	8.208
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.795 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.511 (0.071)	0.600 (0.522, 0.690)			
$South^{*}_{meanday-1to-3}$	-0.054 (0.025)	0.947 (0.903, 0.994)	3245.139	6191.201	7.482
$sin(\frac{2\pi t}{365})$	-0.231 (0.070)	0.794 (0.692, 0.910)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.600 (0.522, 0.690)			
$South^{*}_{meanday-1to-4}$	-0.077 (0.028)	0.926 (0.877, 0.979)	3242.427	6192.267	9.359
$sin(\frac{2\pi t}{365})$	-0.232 (0.070)	0.793 (0.691, 0.909)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.601 (0.523, 0.691)			
$South^{*}_{meanday-1to-5}$	-0.091 (0.031)	0.913 (0.859, 0.971)	3241.296	6192.669	10.200
$sin(\frac{2\pi t}{365})$	-0.232 (0.070)	0.793 (0.691, 0.909)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.601 (0.523, 0.690)			
$South^{*}_{meanday-1to-6}$	-0.089 (0.034)	0.915 (0.857, 0.977)	3243.030	6192.730	7.058
$sin(\frac{2\pi t}{365})$	-0.232 (0.070)	0.793 (0.692, 0.909)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.601 (0.522, 0.690)			
$South^{*}_{meanday-1to-7}$	-0.081 (0.036)	0.922 (0.860, 0.989)	3244.995	6192.426	8.279
$sin(\frac{2\pi t}{365})$	-0.232 (0.070)	0.793 (0.692, 0.910)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.601 (0.522, 0.690)			
$South^{*}_{meanday-1to-8}$	-0.078 (0.038)	0.925 (0.859, 0.996)	3245.991	6191.381	8.409
$sin(\frac{2\pi t}{365})$	-0.236 (0.070)	0.790 (0.689, 0.906)			
$cos(\frac{2\pi t}{365})$	-0.508 (0.071)	0.602 (0.524, 0.692)			
$South^{*}_{meanday-1to-14}$	-0.122 (0.050)	0.885 (0.802, 0.977)	3244.272	6191.702	10.982

Table C.21: Results from logistic regression of Wind Velocity (South), Period 2 ('*' = retained for further analysis).

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.220 (0.070)	0.802 (0.700, 0.920)			
$cos(\frac{2\pi t}{365})$	-0.517 (0.071)	0.597 (0.519, 0.686)			
$East^{d*}_{meanday-3}$	0.038 (0.015)	1.038 (1.008, 1.069)	3244.304	6178.653	4.330
$sin(\frac{2\pi t}{365})$	-0.221 (0.069)	0.802 (0.700, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.517 (0.071)	0.596 (0.518, 0.686)			
$East^{d*}_{maxday-3}$	0.024 (0.009)	1.025 (1.007, 1.042)	3242.501	6183.724	9.806
$sin(\frac{2\pi t}{365})$	-0.221 (0.069)	0.802 (0.700, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.519 (0.072)	0.595 (0.517, 0.685)			
$East^{d*}_{stdday-3}$	0.076 (0.025)	1.079 (1.027, 1.134)	3241.168	6187.624	6.871
$sin(\frac{2\pi t}{365})$	-0.221 (0.069)	0.802 (0.700, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.518 (0.072)	0.596 (0.518, 0.686)			
$East^{d*}_{rangeday-3}$	0.026 (0.009)	1.026 (1.008, 1.045)	3242.306	6186.175	11.493
$sin(\frac{4\pi t}{365})$	-0.223 (0.069)	0.800 (0.699, 0.917)			
$cos(\frac{4\pi t}{365})$	-0.516 (0.071)	0.597 (0.519, 0.687)			
$East^{d}_{maxAHCday-3}$	0.034 (0.015)	1.034 (1.004, 1.065)	3245.367	6191.599	7.937
$sin(\frac{2\pi t}{365})$	-0.221 (0.069)	0.802 (0.700, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.516 (0.071)	0.597 (0.519, 0.687)			
$East^{d}_{meanAHCday-3}$	0.218 (0.066)	1.244 (1.092, 1.416)	3239.689	6180.530	8.431
$sin(\frac{4\pi t}{365})$	-0.223 (0.069)	0.800 (0.699, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.597 (0.519, 0.687)			
$East^{d*}_{stdAHCday-3}$	0.141 (0.059)	1.151 (1.026, 1.292)	3244.589	6187.009	11.700
$sin(\frac{4\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.913)			
$cos(\frac{4\pi t}{365})$	-0.489 (0.072)	0.613 (0.533, 0.706)			
$East^{*}_{maxAHCday-1-2}$	-0.053 (0.012)	0.948 (0.911, 0.986)	3242.910	6193.406	7.446
$sin(\frac{4\pi t}{365})$	-0.228 (0.070)	0.797 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.502 (0.071)	0.605 (0.527, 0.696)			
$East^{*}_{stdAHCday-1-2}$	-0.155 (0.077)	0.857 (0.737, 0.997)	3246.171	6190.341	8.786

Table C.22: Results from logistic regression of Wind Velocity (East), Period 2 ('d' = deseasoned; '*' = retained for further analysis).

Average day effects: Of the eight average day effects examined in relation to SIDS, seven were significant. As previously mentioned, one of the main aims of this section of analysis is data reduction, therefore only two variables are retained for further analysis, namely $South_{meanday-1to-6}$ and $South_{meanday-1to-14}$. The first variable was chosen as it scored best on the goodness-of-fit statistics in comparison to the other average day effects. The second variable was retained as it essentially contains all the information about the mean southerly wind velocity over the two weeks prior to the day of interest, capturing information found in the other average day effect variables. Both variables relate to a decreased risk of SIDS with increased average southern wind velocity over the few days prior to the day of interest.

Between day effects: No significant associations were identified between SIDS incidence and the southern component of wind velocity.

C.2.6 Wind Velocity - East

Significant models from the logistic regression analysis of daily eastern wind velocity are presented in Table C.22. There were some associations between the eastern component of wind velocity and seasonality, with those variables that were deseasoned denoted by a superscripted 'd' in Table C.22. The relationship between eastern wind velocity and SIDS identified similar significant effects as found with southern wind velocity.

Daily summary measures: The two daily summary measures corresponding to the eastern component of wind velocity that were significantly related to the incidence of SIDS, after adjustment for seasonality, were $East_{meanday-3}$ and $East_{maxday-3}$. $East_{meanday-3}$, with lower values on two of the three goodness-of-fit measures as compared to $East_{maxday-3}$, is retained for further analysis. This model showed that a 1 knot increase in $East_{meanday-3}$ corresponded to SIDS risk increasing 1.038 times. This implied that the highest risk of SIDS was in winter, on days where there was a strong easterly component of wind velocity three days prior.

Within day effects: As with southerly wind velocity, all the variables measuring the variation in eastern wind velocity three days before $day0$ were significantly associated with the rate of SIDS, over and above

seasonality. Again, comparison of goodness-of-fit statistics resulted in $East_{meanAHCday-3}$ being retained for further analysis. An increase in the variability of eastern wind velocity resulted in an increased risk of SIDS three days later.

Between day effects: Two between day effects were identified as being related to the incidence of SIDS, namely $East_{maxAHCday-1-2}$ and $East_{stdAHCday-1-2}$. The two measures are essentially capturing the same climatic profiles so only $East_{maxAHCday-1-2}$, with better overall scores on the goodness-of-fit measures, is retained for further analysis. With a deviance statistic of $G = 7.418$ ($p = 0.006$), the addition of $East_{maxAHCday-1-2}$ to the baseline seasonality only model significantly improved the log-likelihood. This model identified a decreased risk of SIDS with increasing $East_{maxAHCday-1-2}$.

Average day effects: No significant associations between the easterly component of wind velocity average day effects and the incidence of SIDS were found.

Between average day effects: As with average day effects, no significant relationships were identified between SIDS and easterly wind velocity.

C.2.7 Wind Velocity - West

The western component of wind velocity did not appear to have an underlying seasonal component, that is, it was not significantly related to seasonality, as measured by a first order sinusoidal function. Models containing those westerly wind velocity variables that were associated with the rate of SIDS, over and above seasonality, are presented in Table C.23.

Daily summary measures: Three variables measuring the daily west wind velocity were significantly related to the incidence of SIDS, namely $West_{minday-4}$, $West_{meanday-3}$ and $West_{maxday-4}$. Utilizing the goodness-of-fit statistics in the decision making process, of the two variables measuring the western component of wind velocity three days prior, $West_{maxday-3}$ is retained for further analysis. An increase in either $West_{minday-4}$, or $West_{maxday-3}$, corresponded to a decreased risk of SIDS.

Within day effects: As with wind velocity from both the south and east, all variables measuring the variation in westerly wind velocity three days before $day0$ were significantly associated with SIDS. In contrast to the other two wind velocity direction variables, $West_{stdday-3}$ is retained for further analysis. This model suggested that the greatest risk of SIDS occurred in winter, on days where there was little variation in the westerly wind velocity three days prior.

Between day effects: Ten between day measures of western wind velocity were significantly related to SIDS incidence, over and above seasonality, as listed in Table C.23. A comparison of goodness-of-fit statistics was made between those competing variables that were essentially capturing the same climatic profile. Variables that are retained for further analysis are $West_{meanday-2-3}$, $West_{meanday-3-4}$, $West_{meanday-7-8}$, $West_{minday0-4}$, $West_{maxday0-1}$ and $West_{rangeday0-1}$. Of these variables, an increase in $West_{meanday-2-3}$, $West_{meanday-3-4}$, or $West_{minday0-4}$ corresponded to a decrease in the risk of SIDS, whereas an increase in $West_{meanday-7-8}$, $West_{maxday0-1}$ or $West_{rangeday0-1}$ related to an increased SIDS risk.

Average day effects: Of the variables measuring the average wind velocity from the west, over the previous lags, four were related to SIDS. $West_{meanday-1to-3}$, with lower D and P statistics than the remaining three candidates is retained for further analysis. The deviance statistic $G = 8.765$ ($p = 0.003$) showed that the inclusion of this variable to the baseline model significantly improved the log-likelihood. An increase in the mean west wind velocity over the three days prior to $day0$ corresponded to a decreased risk of SIDS.

Between average day effects: No significant associations were found between any of the between average day effects and SIDS incidence.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.536 (0.072)	0.585 (0.509, 0.673)			
$West^{*}_{minday-4}$	-0.158 (0.064)	0.854 (0.753, 0.967)	3241.982	6199.913	7.588
$sin(\frac{2\pi t}{365})$	-0.236 (0.070)	0.790 (0.689, 0.905)			
$cos(\frac{2\pi t}{365})$	-0.534 (0.071)	0.587 (0.510, 0.675)			
$West_{meanday-3}$	-0.064 (0.018)	0.938 (0.905, 0.972)	3236.379	6183.503	6.778
$sin(\frac{2\pi t}{365})$	-0.233 (0.070)	0.792 (0.691, 0.908)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.597 (0.520, 0.687)			
$West^{*}_{maxday-3}$	-0.027 (0.008)	0.973 (0.958, 0.989)	3238.801	6183.456	3.195
$sin(\frac{2\pi t}{365})$	-0.232 (0.070)	0.793 (0.692, 0.909)			
$cos(\frac{2\pi t}{365})$	-0.507 (0.071)	0.603 (0.524, 0.693)			
$West^{*}_{stdday-3}$	-0.084 (0.026)	0.920 (0.874, 0.969)	3239.879	6176.460	7.240
$sin(\frac{2\pi t}{365})$	-0.232 (0.070)	0.793 (0.692, 0.909)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.600 (0.522, 0.690)			
$West_{rangeday-3}$	-0.026 (0.008)	0.974 (0.958, 0.990)	3240.099	6180.430	6.606
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.509 (0.071)	0.601 (0.523, 0.691)			
$West_{maxAHCday-3}$	-0.031 (0.012)	0.970 (0.947, 0.993)	3243.623	6184.657	8.544
$sin(\frac{2\pi t}{365})$	-0.231 (0.070)	0.794 (0.693, 0.910)			
$cos(\frac{2\pi t}{365})$	-0.519 (0.071)	0.595 (0.517, 0.684)			
$West_{meanAHCday-3}$	-0.137 (0.049)	0.872 (0.791, 0.960)	3242.191	6177.384	12.029
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.600 (0.522, 0.690)			
$West_{stdAHCday-3}$	-0.106 (0.045)	0.900 (0.824, 0.982)	3244.548	6178.997	10.874
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.535 (0.072)	0.585 (0.509, 0.674)			
$West_{minday-2-3}$	-0.094 (0.046)	0.910 (0.833, 0.995)	3245.204	6202.112	8.042
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.535 (0.072)	0.585 (0.509, 0.674)			
$West_{minday-3-4}$	-0.094 (0.046)	0.910 (0.832, 0.995)	3244.168	6210.731	4.874
$sin(\frac{2\pi t}{365})$	-0.217 (0.070)	0.805 (0.702, 0.923)			
$cos(\frac{2\pi t}{365})$	-0.492 (0.072)	0.611 (0.531, 0.704)			
$West_{minday-7-8}$	0.083 (0.029)	1.065 (1.007, 1.127)	3246.022	6188.237	7.720
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.536 (0.072)	0.585 (0.508, 0.674)			
$West^{*}_{minday0-4}$	-0.073 (0.038)	0.930 (0.864, 1.001)	3245.996	6201.395	6.925
$sin(\frac{2\pi t}{365})$	-0.231 (0.070)	0.794 (0.693, 0.910)			
$cos(\frac{2\pi t}{365})$	-0.518 (0.071)	0.596 (0.518, 0.685)			
$West^{*}_{meanday-2-3}$	-0.061 (0.022)	0.941 (0.901, 0.983)	3242.457	61970978	4.783
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.794 (0.693, 0.910)			
$cos(\frac{2\pi t}{365})$	-0.517 (0.071)	0.596 (0.519, 0.686)			
$West^{*}_{meanday-3-4}$	-0.053 (0.022)	0.948 (0.908, 0.990)	3244.108	6173.692	9.955
$sin(\frac{2\pi t}{365})$	-0.218 (0.070)	0.804 (0.702, 0.922)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.601 (0.523, 0.691)			
$West^{*}_{meanday-7-8}$	0.039 (0.019)	1.040 (1.001, 1.079)	3246.416	6185.593	5.607
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.802 (0.699, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.532 (0.072)	0.587 (0.510, 0.676)			
$West^{*}_{maxday0-1}$	0.023 (0.010)	1.023 (1.004, 1.043)	3245.092	6202.257	4.057
$sin(\frac{2\pi t}{365})$	-0.220 (0.070)	0.803 (0.700, 0.920)			
$cos(\frac{2\pi t}{365})$	-0.532 (0.072)	0.587 (0.510, 0.676)			
$West_{stdday0-1}$	0.060 (0.030)	1.061 (1.000, 1.126)	3246.561	6198.942	8.654
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.802 (0.700, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.533 (0.072)	0.587 (0.510, 0.675)			
$West^{*}_{rangeday0-1}$	0.022 (0.010)	1.023 (1.003, 1.043)	3245.503	6198.907	5.937
$sin(\frac{2\pi t}{365})$	-0.237 (0.070)	0.789 (0.688, 0.904)			
$cos(\frac{2\pi t}{365})$	-0.536 (0.072)	0.585 (0.508, 0.674)			
$West^{*}_{meanAHCday-1to-3}$	-0.071 (0.024)	0.932 (0.889, 0.976)	3240.935	6189.527	4.454
$sin(\frac{2\pi t}{365})$	-0.239 (0.070)	0.788 (0.687, 0.903)			
$cos(\frac{2\pi t}{365})$	-0.538 (0.072)	0.584 (0.507, 0.672)			
$West_{meanAHCday-1to-4}$	-0.076 (0.026)	0.927 (0.880, 0.976)	3241.563	6195.647	3.257
$sin(\frac{2\pi t}{365})$	-0.239 (0.070)	0.787 (0.687, 0.902)			
$cos(\frac{2\pi t}{365})$	-0.539 (0.072)	0.583 (0.507, 0.672)			
$West_{meanAHCday-1to-5}$	-0.078 (0.029)	0.925 (0.875, 0.979)	3242.634	6200.488	8.899
$sin(\frac{2\pi t}{365})$	-0.240 (0.070)	0.787 (0.687, 0.902)			
$cos(\frac{2\pi t}{365})$	-0.539 (0.072)	0.583 (0.506, 0.672)			
$West_{meanAHCday-1to-6}$	-0.078 (0.031)	0.925 (0.871, 0.983)	3243.675	6199.577	8.095

Table C.23: Results from logistic regression of Wind Velocity (West), Period 2 ('*' = retained for further analysis).

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.547 (0.073)	0.579 (0.502, 0.667)			
$Humid^{d*}_{maxday-14}$	-0.014 (0.006)	0.986 (0.974, 0.998)	3248.624	6192.909	8.962
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.492 (0.072)	0.601 (0.531, 0.704)			
$Humid^{*}_{maxAHCday-5}$	-0.012 (0.006)	0.988 (0.976, 1.000)	3246.066	6176.405	7.506
$sin(\frac{2\pi t}{365})$	-0.218 (0.070)	0.804 (0.702, 0.922)			
$cos(\frac{2\pi t}{365})$	-0.502 (0.071)	0.605 (0.526, 0.696)			
$Humid_{minday-5-6}$	-0.010 (0.005)	0.990 (0.980, 0.999)	3245.720	6187.074	14.293
$sin(\frac{2\pi t}{365})$	-0.221 (0.070)	0.802 (0.700, 0.919)			
$cos(\frac{2\pi t}{365})$	-0.505 (0.071)	0.604 (0.525, 0.694)			
$Humid^{*}_{minday0-2}$	-0.009 (0.004)	0.991 (0.983, 1.000)	3246.072	6204.015	9.038
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.492 (0.072)	0.611 (0.531, 0.703)			
$Humid^{*}_{meanday-5-6}$	-0.016 (0.008)	0.984 (0.970, 0.999)	3245.629	6205.701	5.956
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.504 (0.071)	0.604 (0.526, 0.695)			
$Humid^{*}_{rangeday0-3}$	-0.009 (0.004)	0.991 (0.982, 0.999)	3245.656	6169.950	11.512
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.493 (0.071)	0.611 (0.531, 0.703)			
$Humid^{*}_{maxAHCday0-5}$	-0.018 (0.007)	0.982 (0.969, 0.995)	3242.775	6186.947	4.296
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.495 (0.071)	0.610 (0.530, 0.701)			
$Humid_{stdAHCday0-5}$	-0.077 (0.031)	0.926 (0.872, 0.983)	3243.202	6184.249	5.542

Table C.24: Results from logistic regression of Humidity, Period 2 ('d' = deseasoned; '*' = retained for further analysis).

C.2.8 Wind Chill

With the exception of variables measuring the maximum daily wind chill and corresponding lags, all the daily wind chill measures were significantly associated with seasonality. After seasonally adjusting the required variables, wind chill was found to have no relationship with the incidence of SIDS in Period 2.

C.2.9 Relative Humidity

Relative humidity models, with covariates significantly related to SIDS, over and above seasonality, are presented in Table C.24. Of the variables shown, only the maximum humidity a fortnight ago was required to be seasonally adjusted (denoted with 'd'). No relationships were identified between any of the average or between average day effects of humidity and SIDS incidence.

Daily summary measures: $Humid_{maxday-14}$, the maximum humidity fourteen days prior to the day of interest, was the single daily summary measure to be significantly related to the incidence of SIDS. Interpretation of the odds ratio for this model shows that a 1% increase in the maximum relative humidity corresponded to SIDS risk decreasing 0.986 times a fortnight later.

Within day effects: Again, only a single relative humidity variable from this category was related to SIDS. The addition of $Humid_{maxAHCday-5}$ to the baseline seasonality only model significantly improved the log-likelihood ($G = 4.262$, $p = 0.039$). An increase in $Humid_{maxAHCday-5}$ corresponded to a decreased risk of SIDS, implying that the greatest risk of SIDS occurred in winter, on days where there was little variation in relative humidity five days prior.

Between day effects: Six humidity variables measuring between day effects were significantly related to the incidence of SIDS, over and above seasonality, as shown in Table C.24. In choosing between variables with similar climatic profiles, to be retained for further analysis, $Humid_{meanday-5-6}$ was preferred over $Humid_{minday-5-6}$ while $Humid_{meanAHCday0-5}$ was preferred over $Humid_{stdAHCday0-5}$, as both retained variables scored better on two of the three goodness-of-fit statistics than their competitors. The four variables which will be utilized in further analysis of the relationship between humidity between day effects and SIDS are therefore $Humid_{meanday-5-6}$, $Humid_{minday0-2}$, $Humid_{rangeday0-3}$ and $Humid_{meanAHCday0-5}$. An increase in any of these variables corresponded to a decreased risk of SIDS.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.241 (0.070)	0.786 (0.685, 0.903)			
$cos(\frac{2\pi t}{365})$	-0.499 (0.072)	0.607 (0.528, 0.699)			
$Pres^{*}_{minday-4}$	0.009 (0.004)	1.009 (1.001, 1.018)	3245.897	6193.423	7.604
$sin(\frac{2\pi t}{365})$	-0.244 (0.070)	0.783 (0.683, 0.9899			
$cos(\frac{2\pi t}{365})$	-0.495 (0.072)	0.609 (0.530, 0.701)			
$Pres^{*}_{minday-5}$	0.011 (0.004)	1.011 (1.002, 1.020)	3243.987	6198.849	9.521
$sin(\frac{2\pi t}{365})$	-0.241 (0.070)	0.786 (0.685, 0.902)			
$cos(\frac{2\pi t}{365})$	-0.499 (0.072)	0.607 (0.528, 0.699)			
$Pres^{*}_{meanday-4}$	0.009 (0.005)	1.009 (1.000, 1.019)	3246.565	6196.250	8.194
$sin(\frac{2\pi t}{365})$	-0.245 (0.070)	0.783 (0.682, 0.898)			
$cos(\frac{2\pi t}{365})$	-0.495 (0.072)	0.610 (0.530, 0.702)			
$Pres^{*}_{meanday-5}$	-0.012 (0.005)	1.012 (1.002, 1.022)	3244.257	6200.262	10.182
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.547 (0.073)	0.579 (0.502, 0.667)			
$Pres^{*}_{maxday-5}$	-0.014 (0.006)	0.986 (0.974, 0.998)	3244.403	6196.512	14.136
$sin(\frac{2\pi t}{365})$	-0.219 (0.070)	0.803 (0.701, 0.920)			
$cos(\frac{2\pi t}{365})$	-0.506 (0.071)	0.603 (0.524, 0.693)			
$Pres^{*}_{maxday-2-3}$	0.021 (0.011)	1.022 (1.000, 1.044)	3246.743	6187.957	5.202
$sin(\frac{2\pi t}{365})$	-0.226 (0.070)	0.798 (0.696, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.510 (0.071)	0.600 (0.522, 0.690)			
$Pres_{meanAHCday0-8}$	0.597 (0.242)	1.816 (1.130, 2.920)	3244.558	6169.894	6.631
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.516 (0.071)	0.597 (0.519, 0.687)			
$Pres^{*}_{stdAHCday0-7}$	0.526 (0.250)	1.691 (1.036, 2.762)	3246.509	6184.551	17.533
$sin(\frac{2\pi t}{365})$	-0.247 (0.070)	0.781 (0.680, 0.897)			
$cos(\frac{2\pi t}{365})$	-0.494 (0.072)	0.610 (0.530, 0.703)			
$Pres^{*}_{meanday-1to-6}$	0.013 (0.006)	1.013 (1.000, 1.026)	3246.415	6203.530	7.422
$sin(\frac{2\pi t}{365})$	-0.248 (0.071)	0.780 (0.680, 0.896)			
$cos(\frac{2\pi t}{365})$	-0.493 (0.072)	0.611 (0.530, 0.703)			
$Pres^{*}_{meanday-1to-7}$	0.013 (0.007)	1.013 (1.000, 1.027)	3246.445	6204.582	7.010
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.511 (0.071)	0.600 (0.522, 0.690)			
$Pres_{diffday0-5}$	-0.010 (0.005)	0.990 (0.981, 1.000)	3246.619	6187.898	6.356
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.801 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.511 (0.071)	0.600 (0.522, 0.690)			
$Pres_{diffday0-6}$	-0.010 (0.005)	0.991 (0.981, 1.000)	3246.565	6187.228	8.175

Table C.25: Results from logistic regression of Pressure, Period 2 ('*' = retained for further analysis).

C.2.10 Pressure

Pressure was not significantly associated with seasonality. Models containing pressure variables that were significantly associated with the incidence of SIDS, over and above the seasonal component, are presented in Table C.25.

Daily summary measures: Five daily summary measures of pressure were significantly related to SIDS: the minimum and mean pressure four days before *day*0 and the minimum, mean and maximum pressure five days before *day*0. A comparison of goodness-of-fit statistics showed that $Pres_{minday-4}$ is a better measure of the pressure four days before *day*0 than $Pres_{meanday-4}$, and similarly, $Pres_{minday-5}$ was selected as the best measure of the pressure five days before *day*0. Both $Pres_{minday-4}$ and $Pres_{minday-5}$ are therefore retained for further analysis. Interpretation of the models containing either $Pres_{minday-4}$ or $Pres_{minday-5}$ showed that an increase in minimum pressure four, or five, days before *day*0 corresponded to an increase in SIDS risk.

Within day effects: Of the variables measuring the variation in daily pressure, none were significantly related to SIDS incidence, after accounting for seasonality.

Between day effects: A significant association with SIDS was found with three between day effects, namely $Pres_{maxday-2-3}$, $Pres_{meanAHCday0-8}$ and $Pres_{stdAHCday0-7}$. An increase in all of these variables relates to an increased risk of SIDS; specifically, a 1 HPa increase in $Pres_{meanday0-8}$ resulted in SIDS risk increasing 1.816 times.

Average day effects: Two measures of average retrospective pressure were significantly related to SIDS, over and above seasonality: the average pressure over past six days prior to *day*0($Pres_{meanday-1to-6}$) and over the seven days prior to *day*0 ($Pres_{meanday-1to-7}$). Both these variables are capturing very similar

climatic profiles, therefore $Pres_{meanday-1to-6}$, with marginally smaller values on two of the three goodness-of-fit statistics compared to $Pres_{meanday-1to-7}$, is retained for further analysis. This model implied that the greatest SIDS risk occurred in winter, when the pressure was high over the six days prior to the day of interest.

Between average day effects: Two between average pressure day effects were associated with the incidence of SIDS. Inclusion of either $Pres_{diffday0-5}$ or $Pres_{diffday0-6}$ into the seasonality only model significantly improved the log-likelihood ($G = 3.709$ ($p = 0.054$) and $G = 3.763$ ($p = 0.052$) respectively). $Pres_{diffday0-6}$ was retained over $Pres_{diffday0-5}$. This model showed a decreased SIDS risk with increasing $Pres_{diffday0-6}$.

C.2.11 Rainfall

Table C.26 presents models containing rainfall variables found to be significantly related to SIDS, over and above the seasonal component, as measured by a first order sinusoid. Rainfall was not related to seasonality.

Daily summary measures: A single daily summary measure of rainfall was found to be related to the incidence of SIDS, namely $Rain_{maxday0}$, the maximum rainfall in one hour on the day of interest. With a deviance statistic of $G = 6.093$ ($p = 0.014$), there was sufficient evidence to show that the addition of $Rain_{maxday0}$ to the baseline model significantly improved the log-likelihood. Interpretation of the corresponding odds ratio for this model implied that SIDS risk increased 1.080 times with each 1mm increase in $Rain_{maxday0}$.

Within day effects: All the variables measuring the variation in rainfall on the day of interest were significantly related to SIDS incidence, after accounting for seasonality. Of the five variables, $Rain_{stdday0}$, with lower values for D and HL than the other four candidates, is retained for further examination of the relationship between SIDS and the variation in rainfall. This model implied that SIDS risk increased with increasing $Rain_{stdday0}$.

Between day effects: Of all the between day measures examined in relation to SIDS, 27 were found to be significantly associated with the dependent variable, as listed in Table C.26. Those that are retained for further analysis are $Rain_{maxday0-1}$, $Rain_{maxday0-2}$, $Rain_{maxday0-4}$ and $Rain_{maxday0-7}$, $Rain_{stdAHCday0-7}$, $Rain_{stdday0-4}$, $Rain_{rangeday0-7}$, $Rain_{rangeday0-1}$, $Rain_{rangeday0-2}$, $Rain_{maxAHCday0-3}$, $Rain_{stdAHCday0-5}$, $Rain_{maxAHCday0-6}$ and $Rain_{maxAHCday0-14}$. An increase in any of these variables corresponded to an increased risk of SIDS.

Average day effects: The mean rainfall per hour (per day) averaged over the fortnight before $day0$ was the only average daily rainfall effect significantly related to the incidence of SIDS. This model indicates that the highest risk of SIDS occurred in winter, where there had been little rain over the two weeks prior to the day of interest.

Between day effects: Three measures of the difference between rainfall on $day0$ and past averages were related to SIDS: $Rain_{diffday0-2}$, $Rain_{diffday0-3}$ and $Rain_{diffday0-14}$. The first two variables capture very similar climatic profiles and thus only $Rain_{diffday0-3}$, with lower P and HL statistics than $Rain_{diffday0-2}$, will be retained. An increase in either $Rain_{diffday0-3}$ or $Rain_{diffday0-14}$ corresponded to an increase in the risk of SIDS.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.519 (0.071)	0.595 (0.518, 0.684)			
$Rain^{*}_{maxday0}$	0.077 (0.029)	1.080 (1.021, 1.142)	3244.235	6185.284	11.548
$sin(\frac{2\pi t}{365})$	-0.226 (0.070)	0.798 (0.696, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.516 (0.071)	0.597 (0.519, 0.686)			
$Rain_{stdday0}$	0.261 (0.117)	1.298 (1.031, 1.634)	3246.048	6179.358	11.788
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.519 (0.071)	0.595 (0.518, 0.684)			
$Rain_{rangeday0}$	0.077 (0.029)	1.080 (1.022, 1.143)	3244.187	6185.429	9.336

Table C.26: Results from logistic regression of Rainfall, Period 2 ('*' = retained for further analysis).

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.522 (0.071)	0.593 (0.516, 0.682)			
$Rain_{maxAHCday0}$	0.100 (0.034)	1.105 (1.033, 1.182)	3243.365	6189.368	9.257
$sin(\frac{2\pi t}{365})$	-0.226 (0.070)	0.798 (0.696, 0.915)			
$cos(\frac{2\pi t}{365})$	-0.514 (0.071)	0.598 (0.520, 0.688)			
$Rain_{meanAHCday0}$	0.618 (0.259)	1.854 (1.116, 3.081)	3245.368	6178.737	11.857
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.694, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.071)	0.593 (0.515, 0.682)			
$Rain^{*}_{stdAHCday0}$	0.393 (0.129)	1.481 (1.150, 1.907)	3242.745	6190.752	8.979
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.071)	0.592 (0.515, 0.681)			
$Rain^{*}_{maxday0-1}$	0.062 (0.027)	1.064 (1.010, 1.122)	3245.648	6181.050	7.673
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.521 (0.071)	0.594 (0.516, 0.683)			
$Rain^{*}_{maxday0-2}$	0.506 (0.025)	1.052 (1.001, 1.106)	3246.826	6187.517	4.814
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.797 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.071)	0.593 (0.516, 0.682)			
$Rain^{*}_{maxday0-4}$	0.061 (0.024)	1.063 (1.014, 1.115)	3244.793	6192.090	7.029
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.795 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.520 (0.071)	0.594 (0.517, 0.683)			
$Rain^{*}_{maxday0-7}$	0.062 (0.024)	1.063 (1.014, 1.115)	3244.786	6182.603	8.650
$sin(\frac{2\pi t}{365})$	-0.226 (0.070)	0.798 (0.696, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.518 (0.071)	0.596 (0.518, 0.685)			
$Rain^{*}_{stdday0-4}$	0.199 (0.100)	1.223 (1.004, 1.483)	3246.766	6186.078	5.301
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.694, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.517 (0.071)	0.597 (0.519, 0.686)			
$Rain_{stdday0-7}$	0.221 (0.098)	1.248 (1.031, 1.510)	3245.769	6180.029	11.233
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.524 (0.071)	0.592 (0.515, 0.681)			
$Rain^{*}_{rangeday0-1}$	0.063 (0.027)	1.065 (1.010, 1.122)	3245.603	6181.104	8.152
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.521 (0.071)	0.594 (0.516, 0.683)			
$Rain^{*}_{rangeday0-2}$	0.051 (0.026)	1.052 (1.001, 1.106)	3246.781	6187.634	4.846
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.795 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.520 (0.071)	0.594 (0.517, 0.683)			
$Rain^{*}_{rangeday0-7}$	0.062 (0.024)	1.064 (1.014, 1.116)	3244.755	6182.656	8.617
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.795 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.525 (0.072)	0.591 (0.140, 0.680)			
$Rain_{maxAHCday0-1}$	0.073 (0.032)	1.076 (1.010, 1.145)	3245.924	6184.797	13.019
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.795 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.526 (0.072)	0.591 (0.514, 0.680)			
$Rain_{maxAHCday0-2}$	0.074 (0.030)	1.077 (1.015, 1.142)	3245.146	6189.773	6.038
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.795 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.525 (0.071)	0.592 (0.514, 0.681)			
$Rain^{*}_{maxAHCday0-3}$	0.072 (0.030)	1.075 (1.013, 1.140)	3245.427	6186.073	7.147
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.795 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.527 (0.072)	0.590 (0.513, 0.679)			
$Rain_{maxAHCday0-4}$	0.080 (0.029)	1.083 (1.023, 1.147)	3244.027	6195.478	5.877
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.694, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.071)	0.593 (0.515, 0.682)			
$Rain_{maxAHCday0-5}$	0.064 (0.031)	1.066 (1.004, 1.132)	3246.510	6190.458	4.296
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.794 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.525 (0.071)	0.592 (0.514, 0.680)			
$Rain^{*}_{maxAHCday0-6}$	0.075 (0.030)	1.078 (1.016, 1.144)	3245.111	6177.456	12.040
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.795 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.522 (0.071)	0.593 (0.516, 0.682)			
$Rain_{maxAHCday0-7}$	0.070 (0.031)	1.072 (1.010, 1.138)	3245.870	6186.030	10.457
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.794 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.522 (0.071)	0.594 (0.516, 0.683)			
$Rain^{*}_{maxAHCday0-14}$	0.064 (0.031)	1.066 (1.004, 1.133)	3246.565	6188.123	6.311
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.695, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.517 (0.071)	0.596 (0.519, 0.685)			
$Rain_{meanAHCday0-1}$	0.482 (0.250)	1.619 (0.992, 2.643)	3246.984	6177.7375	10.318
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.796 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.527 (0.072)	0.591 (0.513, 0.679)			
$Rain_{stdAHCday0-1}$	0.291 (0.120)	1.338 (1.058, 1.693)	3245.398	6186.153	10.913
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.795 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.527 (0.072)	0.591 (0.513, 0.679)			
$Rain_{stdAHCday0-2}$	0.286 (0.113)	1.3311(1.067, 1.662)	3244.931	6190.301	7.484
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.796 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.525 (0.071)	0.592 (0.515, 0.681)			
$Rain_{stdAHCday0-3}$	0.268 (0.115)	1.308 (1.044, 1.638)	3245.668	6186.823	6.349
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.795 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.528 (0.072)	0.590 (0.513, 0.679)			
$Rain_{stdAHCday0-4}$	0.304 (0.111)	1.355 (1.091, 1.684)	3244.039	6195.329	6.044
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.796 (0.694, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.525 (0.072)	0.591 (0.514, 0.680)			
$Rain^{*}_{stdAHCday0-5}$	0.267 (0.114)	1.306 (1.045, 1.632)	3245.666	6191.945	3.800

Table C.26: Results from logistic regression of Rainfall, Period 2 ('*' = retained for further analysis).

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.514 (0.071)	0.598 (0.520, 0.688)			
$Sun^{d*}_{totalday0}$	-0.027 (0.013)	0.973 (0.949, 0.998)	3245.767	6174.471	16.473
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.698, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.598 (0.520, 0.687)			
$Sun^{d*}_{totalday-3}$	0.027 (0.013)	1.027 (1.001, 1.053)	3246.069	6195.379	7.548
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.801 (0.698, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.553 (0.073)	0.575 (0.499, 0.663)			
$Sun^{*}_{totaldaydiff0-2}$	0.048 (0.015)	1.050 (1.019, 1.081)	3240.350	6185.364	20.081
$sin(\frac{2\pi t}{365})$	-0.225 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.514 (0.071)	0.598 (0.520, 0.687)			
$Sun_{totaldiffday0-2}$	-0.022 (0.011)	0.978 (0.957, 1.000)	3246.378	6187.857	6.864
$sin(\frac{2\pi t}{365})$	-0.225 (0.070)	0.798 (0.696, 0.915)			
$cos(\frac{2\pi t}{365})$	-0.516 (0.071)	0.597 (0.519, 0.687)			
$Sun_{totaldiffday0-3}$	-0.030 (0.012)	0.970 (0.948, 0.992)	3243.418	6184.296	8.970
$sin(\frac{2\pi t}{365})$	-0.226 (0.070)	0.798 (0.696, 0.915)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.598 (0.520, 0.687)			
$Sun_{totaldiffday0-4}$	-0.029 (0.012)	0.972 (0.950, 0.994)	3244.404	6186.447	11.971
$sin(\frac{2\pi t}{365})$	-0.226 (0.070)	0.797 (0.696, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.597 (0.520, 0.687)			
$Sun_{totaldiffday0-5}$	-0.032 (0.012)	0.969 (0.946, 0.992)	3243.161	6184.500	8.079
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.696, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.598 (0.520, 0.687)			
$Sun^{*}_{totaldiffday0-6}$	-0.030 (0.012)	0.970 (0.948, 0.994)	3244.069	6180.830	8.0678
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.695, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.598 (0.520, 0.687)			
$Sun_{totaldiffday0-7}$	-0.029 (0.012)	0.972 (0.949, 0.995)	3244.738	6181.012	7.567
$sin(\frac{2\pi t}{365})$	-0.227 (0.070)	0.797 (0.695, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.514 (0.071)	0.598 (0.520, 0.687)			
$Sun_{totaldiffday0-8}$	-0.028 (0.012)	0.973 (0.950, 0.996)	3245.224	6181.666	12.714
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.794 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.514 (0.071)	0.598 (0.520, 0.688)			
$Sun^{*}_{totaldiffday0-14}$	-0.030 (0.012)	0.971 (0.947, 0.995)	3244.580	6181.031	15.086

Table C.27: Results from logistic regression of Sunshine, Period 2 ('d' = deseasoned; '*' = retained for further analysis).

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.795 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.527 (0.072)	0.591 (0.513, 0.679)			
$Rain_{stdAHCday0-6}$	0.301 (0.113)	1.352 (1.082, 1.688)	3244.367	6178.096	14.069
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.795 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.071)	0.593 (0.515, 0.682)			
$Rain^{*}_{stdAHCday0-7}$	0.282 (0.115)	1.326 (1.059, 1.660)	3245.182	6186.376	7.775
$sin(\frac{2\pi t}{365})$	-0.229 (0.070)	0.796 (0.694, 0.912)			
$cos(\frac{2\pi t}{365})$	-0.521 (0.071)	0.594 (0.516, 0.683)			
$Rain^{*}_{stdAHCday0-8}$	0.236 (0.119)	1.267 (1.003, 1.599)	3246.893	6185.911	4.862
$sin(\frac{2\pi t}{365})$	-0.230 (0.070)	0.794 (0.693, 0.911)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.071)	0.593 (0.515, 0.682)			
$Rain_{stdAHCday0-14}$	0.259 (0.117)	1.295 (1.031, 1.628)	3246.083	6190.593	6.966
$sin(\frac{2\pi t}{365})$	-0.226 (0.071)	0.797 (0.696, 0.914)			
$cos(\frac{2\pi t}{365})$	-0.526 (0.072)	0.591 (0.513, 0.680)			
$Rain^{*}_{meanday-1to-14}$	-1.374 (0.694)	0.253 (0.065, 0.986)	3246.161	6215.503	6.726
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.513 (0.071)	0.599 (0.521, 0.688)			
$Rain_{diffday0-2}$	0.377 (0.167)	1.458 (1.051, 2.021)	3245.572	6174.546	13.667
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.513 (0.071)	0.599 (0.521, 0.688)			
$Rain^{*}_{diffday0-3}$	0.362 (0.167)	1.436 (1.036, 1.992)	3246.007	6173.857	7.526
$sin(\frac{2\pi t}{365})$	-0.225 (0.070)	0.799 (0.697, 0.915)			
$cos(\frac{2\pi t}{365})$	-0.513 (0.071)	0.599 (0.521, 0.688)			
$Rain^{*}_{diffday0-14}$	0.370 (0.169)	1.447 (1.038, 2.017)	3246.122	6175.433	4.582

Table C.26: Results from logistic regression of Rainfall, Period 2 ('*' = retained for further analysis).

C.2.12 Sunshine

Resulting models containing significant sunshine covariates are presented in Table C.27. Some sunshine variables were significantly related to seasonality and thus deseasoned where necessary.

Daily summary measures: Two measures of daily sunshine were significantly associated with the incidence of SIDS, after adjusting for seasonality: the total hours of sunshine on the day of interest, and three days before *day*0. A one hour increase in the total sunshine hours on *day*0 related to SIDS risk decreasing

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.803 (0.700, 0.920)			
$cos(\frac{2\pi t}{365})$	-0.521 (0.072)	0.595 (0.517, 0.685)			
$Rad^{d*}_{totalday0}$	-0.023 (0.010)	0.977 (0.958, 0.997)	3233.326	6168.318	18.625
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.796 (0.694, 0.913)			
$cos(\frac{2\pi t}{365})$	-0.519 (0.072)	0.599 (0.520, 0.689)			
$Rad^{d*}_{totalday0-2}$	0.028 (0.012)	1.030 (1.006, 1.054)	3220.612	6154.195	6.668
$sin(\frac{2\pi t}{365})$	-0.228 (0.070)	0.791 (0.689, 0.907)			
$cos(\frac{2\pi t}{365})$	-0.519 (0.072)	0.598 (0.520, 0.689)			
$Rad_{totaldiffday0-2}$	-0.020 (0.009)	0.980 (0.963, 0.997)	3219.764	6154.356	15.359
$sin(\frac{2\pi t}{365})$	-0.231 (0.070)	0.786 (0.686, 0.902)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.072)	0.596 (0.518, 0.686)			
$Rad_{totaldiffday0-3}$	-0.026 (0.009)	0.974 (0.957, 0.992)	3214.413	6142.569	7.012
$sin(\frac{2\pi t}{365})$	-0.233 (0.070)	0.784 (0.684, 0.900)			
$cos(\frac{2\pi t}{365})$	-0.522 (0.072)	0.597 (0.519, 0.687)			
$Rad_{totaldiffday0-4}$	-0.024 (0.009)	0.976 (0.958, 0.994)	3213.697	6130.787	8.502
$sin(\frac{2\pi t}{365})$	-0.236 (0.070)	0.789 (0.687, 0.905)			
$cos(\frac{2\pi t}{365})$	-0.523 (0.072)	0.593 (0.514, 0.683)			
$Rad_{totaldiffday0-5}$	-0.026 (0.010)	0.974 (0.956, 0.992)	3195.758	6123.228	7.094
$sin(\frac{2\pi t}{365})$	-0.232 (0.070)	0.787 (0.685, 0.903)			
$cos(\frac{2\pi t}{365})$	-0.522 (0.072)	0.594 (0.515, 0.684)			
$Rad_{totaldiffday0-6}$	-0.025 (0.010)	0.975 (0.957, 0.993)	3194.733	6106.196	8.5608
$sin(\frac{2\pi t}{365})$	-0.238 (0.070)	0.785 (0.684, 0.901)			
$cos(\frac{2\pi t}{365})$	-0.520 (0.072)	0.594 (0.516, 0.684)			
$Rad_{totaldiffday0-7}$	-0.023 (0.010)	0.976 (0.958, 0.995)	3193.826	6092.986	12.116
$sin(\frac{2\pi t}{365})$	-0.240 (0.070)	0.786 (0.685, 0.903)			
$cos(\frac{2\pi t}{365})$	-0.519 (0.072)	0.593 (0.515, 0.683)			
$Rad^{*}_{totaldiffday0-8}$	-0.022 (0.010)	0.977 (0.958, 0.996)	3187.279	6082.928	11.137

Table C.28: Results from logistic regression of Solar Radiation, Period 2 ('d' = deseasoned; '*' = retained for further analysis).

0.973 times, whereas a one hour increase in the total sunshine hours three days prior corresponded to a 1.027 times increase in the risk of SIDS.

Between day effects: The difference in the total hours of sunshine on the day of interest and two days prior was the only sunshine variable measuring between day effects that was significantly related to the incidence of SIDS, over and above the seasonal component. This model implied SIDS risk increased with increasing $Sun_{totalday0-2}$.

Average day effects: There was no evidence of any average sunshine day effects having any relationship with the rate of SIDS.

Between average day effects: All the variables measuring the difference between the total hours of sunshine on the day of interest and past averages were significantly related to the incidence of SIDS. Of these eight variables, two are retained for further analysis, namely $Sun_{totaldiffday0-6}$ and $Sun_{totaldiffday0-14}$: the first variable is retained as it scored best out of the between average day effects in terms of the goodness-of-fit statistics, while the second was chosen as it captures information about the sunshine over the two weeks prior to $day0$, essentially incorporating the information measured by the other variables. Both $Sun_{totaldiffday0-6}$ and $Sun_{totaldiffday0-14}$ were associated with decreased SIDS risk, implying that the greatest risk of SIDS occurred in winter on days where there had been less sunshine than there was on average over the week or two before $day0$.

C.2.13 Solar Radiation

Table C.28 presents results from modelling the incidence of SIDS with respect to solar radiation, with only models containing variables significantly related to SIDS given. As with sunshine, some of the solar radiation variables contained a significant seasonal component so were therefore adjusted accordingly.

Daily summary measures: The total global solar radiation on the day of interest was the only variable that was identified as significantly related to SIDS, after adjusting for seasonality. The odds ratio for this model implied a 0.977 decrease in the risk of SIDS with every 1 Mj/m^2 increase in $Rad_{totalday0}$.

Between day effects: Again, only a single solar radiation between day effect measure was significantly associated with the incidence of SIDS, namely $Rad_{totalday0-2}$, the difference in solar radiation on the day of interest and two days prior. This model indicated that an increase in $Rad_{totalday0-2}$ corresponded to an increased SIDS risk.

Average day effects: No relationship between average day effects of solar radiation and SIDS incidence was identified, after adjustment for seasonality.

Between average day effects: With the exception of $Rad_{totaldiffday0-14}$, all the solar radiation between average day effects were significantly related to SIDS, over and above the seasonal component. $Rad_{totaldiffday0-8}$, with lower scores on both the D and P goodness-of-fit measures, is retained for further examination of the relationship between SIDS and the difference in solar radiation from past averages. This model showed that an increase in $Rad_{totaldiffday0-8}$ resulted in a decrease in the risk of SIDS.

C.2.14 Dewpoint

Dewpoint showed no significant associations with the seasonality measure. Table C.29 presents dewpoint models corresponding to variables found to be significantly related to the incidence of SIDS.

Daily summary measures: Four daily summary measures of dewpoint were found to be significantly related to SIDS, namely $Dew_{meanday-1}$, $Dew_{meanday-2}$, $Dew_{maxday-1}$ and $Dew_{maxday-2}$. Of the competing variables, on the basis of goodness-of-fit statistics, $Dew_{maxday-1}$ was preferred over $Dew_{meanday-1}$, whereas $Dew_{meanday-2}$ was preferred over $Dew_{maxday-2}$. Both $Dew_{maxday-1}$ and $Dew_{meanday-2}$ are retained for further analysis. These models showed that an increase in either the maximum dewpoint on the day prior to the day of interest, or the mean dewpoint two days before the day of interest, corresponded to a subsequent decrease in SIDS risk.

Within day effects: Two variables measuring the variation in dewpoint on $day0$ were related to SIDS, over and above the seasonal component. The addition of either $Dew_{meanAHCday0}$ or $Dew_{stdAHCday0}$ to the baseline model significantly increased the log-likelihood ($G = 5.759$ ($p = 0.016$) and $G = 4.849$ ($p = 0.028$) respectively). $Dew_{stdAHCday0}$, with lower P and HL statistics, is retained. The odds ratio for this model implied a decreased SIDS risk by 0.734 times for every 1°C increase in $Dew_{stdAHCday0}$.

Between day effects: Of the nine significant between day dewpoint effects presented in Table C.29, seven are retained for further analysis; two pairs of candidate covariates are capturing similar climatic profiles thus $Dew_{stdAHCday-2-3}$ is retained over $Dew_{meanAHCday-2-3}$ and $Dew_{meanAHCday0-8}$ is retained over $Dew_{stdAHCday0-8}$. An increase in any of the between day dewpoint measures related to a decreased risk of SIDS, with the exception of $Dew_{maxday0-14}$.

Average day effects: Both the average dewpoint over the two days prior to $day0$and the average dewpoint over the three days prior to $day0$ were identified as being significantly associated with the incidence of SIDS, over and above seasonality. With lower D and P goodness-of-fit statistics, $Dew_{meanday-1to-2}$ is selected over $Dew_{meanday-1to-3}$ for further analysis. This model highlights an increase in $Dew_{meanday-1to-2}$ leading to a decreased risk of SIDS, implying that the greatest risk of SIDS occurred in winter, on days where the dewpoint was less than it had been over the two days prior to the day of interest.

Variable	Estimate (SE)	Odds ratio (95% CI)	Deviance	Pearson	Hosmer-Lemeshow
$sin(\frac{2\pi t}{365})$	-0.216 (0.070)	0.805 (0.703, 0.923)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Dew_{meanday-1}$	-0.025 (0.013)	0.976 (0.952, 1.000)	3246.422	6188.884	9.410
$sin(\frac{2\pi t}{365})$	-0.216 (0.070)	0.806 (0.599, 0.974)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Dew^{*}_{meanday-2}$	-0.026 (0.013)	0.974 (0.951, 0.999)	3245.986	6191.951	9.961
$sin(\frac{2\pi t}{365})$	-0.217 (0.070)	0.805 (0.702, 0.923)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.600 (0.521, 0.689)			
$Dew^{*}_{maxday-1}$	-0.028 (0.013)	0.973 (0.948, 0.998)	3245.932	6187.121	6.975
$sin(\frac{2\pi t}{365})$	-0.217 (0.070)	0.805 (0.702, 0.923)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Dew_{maxday-2}$	-0.027 (0.013)	0.974 (0.949, 0.999)	3246.188	6192.570	8.480
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Dew_{meanAHCday0}$	-0.365 (0.154)	0.694 (0.513, 0.939)	3244.569	6192.134	4.324
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.511 (0.071)	0.600 (0.522, 0.690)			
$Dew^{*}_{stdAHCday0}$	-0.309 (0.145)	0.734 (0.552, 0.976)	3245.479	6191.834	4.056
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.800 (0.698, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.509 (0.071)	0.601 (0.523, 0.691)			
$Dew^{*}_{maxday0-4}$	-0.033 (0.024)	0.967 (0.923, 1.014)	3243.726	6191.730	4.535
$sin(\frac{2\pi t}{365})$	-0.225 (0.070)	0.798 (0.697, 0.915)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Dew^{*}_{maxday0-7}$	-0.035 (0.023)	0.966 (0.923, 1.011)	3244.300	6183.171	6.485
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.514 (0.071)	0.598 (0.520, 0.687)			
$Dew^{*}_{stdday0-14}$	0.127 (0.062)	1.135 (1.006, 1.281)	3246.243	6183.675	7.084
$sin(\frac{2\pi t}{365})$	-0.222 (0.070)	0.801 (0.699, 0.918)			
$cos(\frac{2\pi t}{365})$	-0.515 (0.071)	0.598 (0.520, 0.687)			
$Dew_{meanAHCday-2-3}$	-0.437 (0.215)	0.646 (0.424, 0.983)	3245.985	6214.096	3.839
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.511 (0.071)	0.600 (0.522, 0.690)			
$Dew^{*}_{meanAHCday0-8}$	-0.487 (0.189)	0.614 (0.425, 0.889)	3243.260	6189.196	4.245
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Dew^{*}_{stdAHCday-2-3}$	-0.404 (0.165)	0.668 (0.484, 0.922)	3243.656	6202.478	4.173
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.600 (0.522, 0.689)			
$Dew^{*}_{stdAHCday0-2}$	-0.360 (0.157)	0.698 (0.513, 0.949)	3244.555	6172.129	11.131
$sin(\frac{2\pi t}{365})$	-0.224 (0.070)	0.799 (0.697, 0.916)			
$cos(\frac{2\pi t}{365})$	-0.511 (0.071)	0.600 (0.522, 0.690)			
$Dew^{*}_{stdAHCday0-3}$	-0.385 (0.157)	0.680 (0.500, 0.925)	3243.718	6202.432	11.296
$sin(\frac{2\pi t}{365})$	-0.223 (0.070)	0.800 (0.698, 0.917)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Dew_{stdAHCday0-8}$	-0.317 (0.150)	0.729 (0.544, 0.977)	3245.473	6175.481	10.645
$sin(\frac{2\pi t}{365})$	-0.216 (0.070)	0.806 (0.703, 0.924)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Dew^{*}_{meanday-1to-2}$	-0.028 (0.013)	0.973 (0.948, 0.998)	3245.863	6190.547	10.830
$sin(\frac{2\pi t}{365})$	-0.216 (0.070)	0.806 (0.703, 0.924)			
$cos(\frac{2\pi t}{365})$	-0.512 (0.071)	0.599 (0.521, 0.689)			
$Dew_{meanday-1to-3}$	-0.027 (0.014)	0.974 (0.948, 1.000)	3246.427	6192.692	10.564

Table C.29: Results from logistic regression of Dewpoint, Period 2 ('*' = retained for further analysis).

Variable	Estimate (SE)	Odds Ratio (95% CI)
Intercept	0.71 (0.36)	
$sin(\frac{2\pi t}{236})$	0.23 (0.07)	1.26 (1.09, 1.45)
$cos(\frac{2\pi t}{365})$	0.49 (0.08)	1.64 (1.40, 1.92)
$East_{maxAHCday-1-2}$	0.06 (0.02)	1.07 (1.02, 1.11)
$North_{meanAHCday0-4}$	-0.18 (0.05)	0.84 (0.76, 0.93)
$North_{maxday-7}$	0.03 (0.01)	1.03 (1.01, 1.05)
$South_{meanday-1to-14}$	0.16 (0.06)	1.18 (1.06, 1.31)
$South_{maxday0-5}$	0.02 (0.01)	1.02 (1.00, 1.04)
$South_{minday-14}$	-0.15 (0.05)	0.86 (0.79, 0.94)
$West_{maxAHCday-1}$	0.03 (0.01)	1.03 (1.00, 1.05)
$West_{maxday0-1}$	-0.03 (0.01)	0.97 (0.95, 0.99)
$Dew_{meanAHCday0-8}$	0.50 (0.19)	1.66 (1.13, 2.42)
$Dew_{stdAHCday-2-3}$	0.40 (0.17)	1.49 (1.07, 2.08)
$Dew_{stdAHCday-5-6}$	-0.29 (0.13)	0.75 (0.58, 0.98)
$East^{d}_{meanAHCday-3}$	-0.21 (0.07)	0.81 (0.70, 0.93)
$Humid^{d}_{maxday-14}$	0.01 (0.01)	1.01 (1.00, 1.03)
$Humid^{d}_{minday-8}$	0.01 (0.00)	1.01 (1.00, 1.01)
$Dew_{maxday0-7}$	0.06 (0.03)	1.07 (1.01, 1.12)
$Dew_{stdday0-14}$	-0.15 (0.06)	0.87 (0.76, 0.98)
$Humid_{maxAHCday-5}$	0.02 (0.01)	1.02 (1.00, 1.03)
$Humid_{maxAHCday-2-3}$	-0.02 (0.01)	0.98 (0.97, 1.00)
$Humid_{minday0-2}$	0.02 (0.01)	1.02 (1.01, 1.03)
$Humid_{rangeday0-3}$	0.01 (0.00)	1.01 (1.00, 1.02)
$Pres_{stdAHCday0-6}$	-0.57 (0.27)	0.56 (0.33, 0.95)
$Pres_{stdAHCday-4-5}$	-0.76 (0.29)	0.47 (0.27, 0.82)
$Rain_{maxday0-4}$	-0.37 (0.15)	0.69 (0.52, 0.92)
$Rain_{stdday0-4}$	1.25 (0.60)	3.49 (1.07, 11.41)
$Sun_{totalday0-2}$	-0.05 (0.02)	0.95 (0.92, 0.98)
$Temp_{meanAHCday0}$	0.42 (0.14)	1.53 (1.16, 2.00)
$Temp_{meanAHCday0-2}$	-0.44 (0.19)	0.64 (0.44, 0.93)
$Temp_{meanday0-14}$	-0.05 (0.02)	0.96 (0.92, 1.00)
$Temp_{minday0-7}$	0.06 (0.02)	1.06 (1.02, 1.10)
$Temp_{rangeday0-5}$	-0.04 (0.01)	0.96 (0.93, 0.99)
$WindD_{maxAHCday-3}$	0.00 (0.00)	1.00 (1.00, 1.01)
$WindD_{maxAHCday-4-5}$	0.00 (0.00)	1.00 (1.00, 1.01)
$WindS_{meanAHCday-5-6}$	0.19 (0.08)	1.21 (1.04, 1.42)
$WindS_{meanday-7-8}$	-0.05 (0.02)	0.95 (0.92, 0.99)
$WindS_{stdday0}$	0.10 (0.03)	1.11 (1.04, 1.18)

Table C.30: Results from multiple logistic regression of daily climate variables, Period 2.

C.2.15 Multiple Logistic Regression Models

Table C.30 shows the results from a multiple logistic model. All 171 candidate climate variables were initially included in the model, and as described in the Period 1 results section, a backwards elimination procedure used to give the resulting model. This model again highlights the multi-factorial relationship between SIDS and climate; with the exception of radiation and wind chill, the model incorporates variants from the complete suite of available climate variables. With 35 climate-based regressors remaining in the multiple logistic regression in Period 2, interpreting this model in terms of a risk profile of SIDS incidence is not practical. As in Period 2, principal component regression (PCR, see Chapter 4 for details) is therefore used to further examine the SIDS-climate daily relationship in Period 2.

Appendix D

Principal Component Regression

This appendix details a full list of climate variables found to be significantly related to the incidence of SIDS in daily univariate analysis in Chapter 3, at an $\alpha \leqslant 0.10$ level of significance. Also presented are those climate variables that significantly contribute to the makeup of various principal components calculated in Chapter 4.

$Temp_{minday0}$	$North_{meanday0-3}$	$West_{stdAHCday0-8}$	$Dew_{minday-4}$
$Temp_{stdday-7}$	$North_{maxday0-6}$	$West_{meanAHCday-4-5}$	$Dew_{minday-6}$
$Temp_{meanday-6-7}$	$North_{maxday-3-4}$	$West_{meanday-1to-2}$	$Dew_{maxday-5}$
$Temp_{meanday0-2}$	$North_{maxday0-5}$		$Dew_{meanday-7}$
$Temp_{maxday-7-8}$	$North_{maxAHCday-3-4}$	$Humid_{stdday-6-7}$	$Dew_{maxday-8}$
$Temp_{rangeday-6-7}$	$North_{rangeday0-5}$	$Humid_{stdAHCday0-5}$	$Dew_{stdday-1}$
$Temp_{stdday0-5}$	$North_{rangeday0-6}$	$Humid_{stdAHCday-4-5}$	$Dew_{meanday-1-2}$
$Temp_{meanAHCday-1-2}$	$North_{maxAHCday-7-8}$		$Dew_{meanday0-2}$
$Temp_{diffday0-7}$	$North_{diffday0-4}$	$Pres_{rangeday0}$	$Dew_{maxday-6-7}$
$Temp_{diffday0-14}$	$North_{diffday0-14}$	$Pres_{stdday-3}$	$Dew_{maxday0-3}$
		$Pres_{rangeday-14}$	$Dew_{maxday-2-3}$
$WindS_{meanday-1}$	$South_{meanday-7}$	$Pres_{meanAHCday-7}$	$Dew_{maxday0-14}$
$WindS_{meanday-2}$	$South_{stdday-2}$	$Pres_{maxAHCday-8}$	$Dew_{stdday-2-3}$
$WindS_{maxday-7}$	$South_{meanAHCday-7}$	$Pres_{meanday-5-6}$	$Dew_{stdday0-4}$
$WindS_{maxday-8}$	$South_{maxAHCday-14}$	$Pres_{maxay-7-8}$	$Dew_{rangeday0-2}$
$WindS_{rangeday-1}$	$South_{maxday0-5}$	$Pres_{rangeday-6-7}$	$Dew_{maxAHCday0-5}$
$WindS_{rangeday-2}$	$South_{stdday0-5}$	$Pres_{stdday0-5}$	$Dew_{meanAHCday-1-2}$
$WindS_{maxAHCday-8}$	$South_{stdday-2-3}$	$Pres_{maxAHCday-7-8}$	$Dew_{meanday-1to-8}$
$WindS_{stdday-14}$	$South_{maxAHCday0-4}$	$Pres_{meanAHCday-1-2}$	$Dew_{meanday-1-14}$
$WindS_{meanday-1-2}$	$South_{meanday-1to-2}$	$Pres_{meanAHCday-2-3}$	$Dew_{diffday0-8}$
$WindS_{maxday0-5}$			$Dew_{diffday0to-14}$
$WindS_{meanAHCday-4-5}$	$East_{meanAHCday0}$	$Rain_{maxday-1}$	
$WindS_{maxAHCday-7-8}$	$East_{meanAHCday-8}$	$Rain_{totalday-7}$	$WindD_{maxAHCday-4}$
$WindS_{meanday-1to-3}$	$East_{minnday-3-4}$	$Rain_{stdday-1}$	$WindD_{stdAHCday-1-2}$
$WindS_{meanday-1to-14}$	$East_{stdAHCday-3-4}$	$Rain_{stdday-7}$	$WindD_{meanAHCday-2-3}$
$WindS_{diffday0-2}$	$East_{meanAHCday0-6}$	$Rain_{stdday-14}$	
$WindS_{diffday0to-14}$		$Rain_{totalday0-5}$	$WindC_{meanday0}$
	$West_{meanday-1}$	$Rain_{totalday0-7}$	$WindC_{meanday-2}$
$North_{minday-7}$	$West_{meanday-7}$	$Rain_{maxday0-1}$	$WindC_{rangeday0}$
$North_{maxday0}$	$West_{maxday-2}$	$Rain_{stdday0-1}$	$WindC_{meanAHCday-7}$
$North_{maxday-8}$	$West_{rangeday-2}$	$Rain_{meanAHCday0-5}$	
$North_{stdday0}$	$West_{stdAHCday-8}$	$Rain_{stdAHCday-6-7}$	
$North_{stdday-8}$	$West_{meanday0-1}$		
$North_{minday0-1}$	$West_{maxday0-3}$	$Sun_{totalday-2-3}$	
$North_{minday-6-7}$	$West_{stdday0-1}$	$Sun_{totalday0-7}$	
$North_{minday0-7}$	$West_{stdday0-3}$		

Table D.1: Climate variables significantly related to SIDS incidence, Period1 ($\alpha \leqslant 0.10$).

$Temp_{minday-8}$	$South_{maxday-5}$	$Humid_{meanAHCday-5-6}$	$Sun_{totalday0-2}$
$Temp_{stdday-3}$	$South_{stdday-3}$	$Humid_{maxAHCday-2-3}$	$Sun_{totalday0-4}$
$Temp_{stdAHCday-8}$	$South_{meanday-3-4}$	$Humid_{maxAHCday0-5}$	$Sun_{diffday0-6}$
$Temp_{meanAHCday0}$	$South_{minday0-14}$	$Humid_{meanAHCday-4-5}$	$Sun_{diffday0-14}$
$Temp_{minday0-7}$	$South_{meanday-4-5}$	$Humid_{meanAHCday-6-7}$	
$Temp_{meanday0-3}$	$South_{maxday0-5}$	$Humid_{meanAHCday0-6}$	$Rad_{totalday0}$
$Temp_{meanday0-14}$	$South_{rangeday0-5}$		$Rad_{totalday-3}$
$Temp_{rangeday0-5}$	$South_{maxAHCday-7-8}$	$Pres_{minday-4}$	$Rad_{totalday0-2}$
$Temp_{maxAHCday-2-3}$	$South_{meanday-1to-6}$	$Pres_{minday-5}$	$Rad_{diffday0-8}$
$Temp_{maxAHCday0-8}$	$South_{meanday-1to-14}$	$Pres_{maxday-6}$	$Rad_{diffday0-14}$
$Temp_{meanAHCday0-2}$	$South_{diffday0-4}$	$Pres_{maxday-2-3}$	
		$Pres_{maxday-4-5}$	$Dew_{maxday-1}$
$WindS_{minday-8}$	$East_{minday-4}$	$Pres_{maxday0-8}$	$Dew_{meanday-2}$
$WindS_{meanday-14}$	$East_{meanday-3}$	$Pres_{meanAHCday0-8}$	$Dew_{maxday0}$
$WindS_{maxday0}$	$East_{meanday-7}$	$Pres_{stdAHCday-4-5}$	$Dew_{maxday-4}$
$WindS_{stdday0}$	$East_{meanAHCday-3}$	$Pres_{stdAHCday0-6}$	$Dew_{stdAHCday0}$
$WindS_{maxAHCday-3}$	$East_{minday-3-4}$	$Pres_{stdAHCday0-7}$	$Dew_{meanday0-1}$
$WindS_{stdAHCday-5}$	$East_{maxAHCday-1-2}$	$Pres_{meanday-1to-5}$	$Dew_{maxday-3-4}$
$WindS_{minday-5-6}$	$East_{meanday-1to-3}$	$Pres_{meanday-1to-14}$	$Dew_{maxday-1}$
$WindS_{meanday-7-8}$		$Pres_{diffday0-6}$	$Dew_{meanday0-8}$
$WindS_{minday0-8}$	$West_{maxday-3}$		$Dew_{maxday0-4}$
$WindS_{maxday0-5}$	$West_{minday-4}$	$Rain_{maxday0}$	$Dew_{maxday0-6}$
$WindS_{rangeday0-5}$	$West_{minday-6}$	$Rain_{stdAHCday0}$	$Dew_{maxday0-7}$
$WindS_{meanAHCday0-8}$	$West_{maxday-2}$	$Rain_{maxday0-1}$	$Dew_{maxday0-14}$
$WindS_{meanAHCday-5-6}$	$West_{stdday-3}$	$Rain_{maxday0-2}$	$Dew_{stdday0-14}$
$WindS_{maxAHCday0-3}$	$West_{maxAHCday-1}$	$Rain_{maxday0-3}$	$Dew_{maxAHCday-1-2}$
$WindS_{maxAHCday0-6}$	$West_{minday-1-2}$	$Rain_{maxday0-4}$	$Dew_{stdAHCday0-3}$
	$West_{meanday-2-3}$	$Rain_{maxday0-5}$	$Dew_{stdAHCday-2-3}$
$North_{minday-14}$	$West_{meanday-3-4}$	$Rain_{maxday0-6}$	$Dew_{meanAHCday-7-8}$
$North_{meanday0}$	$West_{minday-4-5}$	$Rain_{maxday0-7}$	$Dew_{meanAHCday0-8}$
$North_{maxday-7}$	$West_{meanday-7-8}$	$Rain_{rangeday0-1}$	$Dew_{stdAHCday-5-6}$
$North_{rangeday-7}$	$West_{minday0-4}$	$Rain_{stdday0-4}$	$Dew_{stdAHCday0-2}$
$North_{meanday0-14}$	$West_{maxday0-1}$	$Rain_{rangeday0-7}$	$Dew_{meanday-1to-2}$
$North_{meanday-5-6}$	$West_{rangeday0-1}$	$Rain_{rangeday0-2}$	
$North_{meanday0-3}$	$West_{stdday0-3}$	$Rain_{maxAHCday0-3}$	$WindD_{stdAHCday0}$
$North_{meanday0-8}$	$West_{meanday-1to-3}$	$Rain_{maxAHCday0-4}$	$WindD_{maxAHCday-3}$
$North_{maxday-2-3}$	$West_{diffday0-3}$	$Rain_{maxAHCday0-5}$	$WindD_{maxAHCday-4-5}$
$North_{stdAHCday-2-3}$		$Rain_{maxAHCday0-6}$	$WindD_{stdAHCday0-3}$
$North_{stdday0-8}$	$Humid_{minday-1}$	$Rain_{maxAHCday0-8}$	$WindD_{stdAHCday0-6}$
$North_{meanAHCday0-4}$	$Humid_{minday-8}$	$Rain_{maxAHCday0-14}$	
$North_{meanAHCday0-14}$	$Humid_{maxday-14}$	$Rain_{meanday-1to-14}$	$WindC_{maxAHCday-2}$
$North_{diffday0-2}$	$Humid_{maxAHCday-5}$	$Rain_{diffday0-3}$	$WindC_{stdAHCday-4}$
	$Humid_{meanAHCday-8}$	$Rain_{diffday0-14}$	$WindC_{stdAHCday-6}$
$South_{minday-2}$	$Humid_{meanday-5-6}$		
$South_{meanday-4}$	$Humid_{minday0-2}$	$Sun_{totalday0}$	
$South_{minday-14}$	$Humid_{meanday-6-7}$	$Sun_{totalday-3}$	
$South_{maxday-3}$	$Humid_{rangeday0-3}$	$Sun_{totalday-5}$	

Table D.2: Climate variables significnatly related to SIDS incidence, Period 2 ($\alpha \leqslant 0.10$).

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
$Temp_{maxday-7-8}$					0.354								
$Temp_{diffday0-7}$		0.650											
$Temp_{diffday0-14}$		0.680											
$Temp_{minday0}$		0.400											
$WindS_{meanday-1}$				0.698									
$WindS_{meanday-2}$				0.733									
$WindS_{maxday-7}$									0.738				
$WindS_{maxday-8}$									0.839				
$WindS_{rangeday-1}$				0.698						0.465			
$WindS_{rangeday-2}$				0.793									
$WindS_{maxAHCday-8}$									0.692				
$WindS_{meanday-1-2}$				0.343									
$WindS_{maxday0-5}$										0.573			
$WindS_{maxAHCday-7-8}$									0.524				
$WindS_{meanday-1to-3}$				0.791									
$WindS_{meanday-1to-14}$	0.316			0.444					0.453				
$WindS_{diffday0-2}$				-0.511						0.634			
$WindS_{diffday0-14}$										0.706			
$North_{maxday0}$		0.875											
$North_{maxday-8}$					0.916								
$North_{stdday0}$		0.852											
$North_{stdday-8}$					0.893								
$North_{meanday0-3}$		0.441											
$North_{maxday-50}$							0.811						
$North_{maxday0-6}$		0.332					0.768						
$North_{rangeday0-5}$							0.810						
$North_{rangeday0-6}$		0.330					0.761						
$North_{maxAHCday-7-8}$					0.445								
$North_{diffday0-4}$		0.688											
$North_{diffday0-14}$		0.797											
$South_{meanday-7}$						0.333							
$South_{stdday-2}$			0.532										
$South_{meanAHCday-7}$						0.304							
$South_{maxday0-5}$								0.877					
$South_{stdday0-5}$								0.829					
$South_{maxAHC0-4}$								0.364					
$South_{meanday-1to-2}$			0.476										
$East_{meanAHCday0}$		0.393											
$East_{meanAHCday8}$					0.367								
$West_{meanday-1}$			0.404										
$West_{maxday-2}$			0.980										
$West_{rangeday-2}$			0.975										
$West_{stdAHCday-8}$					0.409								
$West_{maxday0-3}$		0.422						0.448					
$West_{stdday0-3}$		0.439						0.398					
$West_{stdAHCday0-8}$					0.310								
$West_{meanday-1to2}$			0.741										
$Pres_{rangeday0}$								0.316					
$Pres_{maxAHCday8}$					0.396								
$Pres_{meanAHCday7}$					0.417								
$Pres_{maxday-7-8}$					0.379								
$Rain_{totalday-7}$						0.710							
$Rain_{stdday-7}$						0.643							
$Rain_{totalday0-5}$						0.608		0.325					
$Rain_{totalday0-7}$						0.923							
$Rain_{maxday0-1}$						0.326							
$Rain_{stdday0-1}$						0.355							
$Rain_{meanAHCday0-5}$						0.507							
$Rain_{stdAHCday-6-7}$						0.314							
$Dew_{meanday-7}$	0.886												
$Dew_{minday-4}$	0.824												
$Dew_{minday-6}$	0.875												
$Dew_{maxday-5}$	0.892												
$Dew_{maxday-8}$	0.846												
$Dew_{meanday-1to-8}$	0.971												
$Dew_{meanday-1to-14}$	0.944												
$WindD_{maxAHCday-4}$											0.985		
$WindD_{meanAHCday-2-3}$													0.988
$WindD_{stdAHCday-1-2}$												0.992	

Table D.3: Component loadings for 13 rotated components, Period 1.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
$Temp_{stdday-7}$				-0.326									
$Temp_{maxday-7-8}$					0.312								
$Temp_{diffday0-7}$		0.707											
$Temp_{diffday0-14}$		0.737											
$Temp_{minday0}$		0.475											
$WindS_{meanday-1}$								0.761					
$WindS_{meanday-2}$						0.804							
$WindS_{maxday-8}$													0.950
$WindS_{rangeday-1}$								0.942					
$WindS_{rangeday-2}$						0.854							
$WindS_{maxAHCday-8}$													0.769
$WindS_{maxAHCday-7-8}$													0.383
$WindS_{meanday-1to-3}$						0.628		0.479					
$WindS_{meanday-1to-14}$	0.329					0.350							0.317
$WindS_{diffday0-2}$						-0.671							
$WindS_{diffday0-14}$						-0.375		0.323					
$North_{maxday0}$		0.915											
$North_{maxday-8}$					0.971								
$North_{stdday0}$		0.881											
$North_{stdday-8}$					0.945								
$North_{meanday0-3}$		0.412											
$North_{maxday-3-4}$										0.953			
$North_{maxday0-6}$									0.927				
$North_{rangeday0-6}$									0.930				
$North_{maxAHCday-3-4}$										0.831			
$North_{maxAHCday-7-8}$					0.440								
$North_{diffday0-4}$		0.742											
$North_{diffday0-14}$		0.847											
$South_{meanday-7}$				0.582									
$South_{stdday-2}$			0.540										
$South_{meanAHCday-7}$				0.528									
$South_{maxday0-5}$											0.981		
$South_{stdday0-5}$											0.936		
$South_{meanday-1to-2}$			0.490										
$East^{*}_{meanAHCday0}$		0.459											
$East^{*}_{meanAHCday8}$					0.391								
$West_{meanday-1}$			0.404										
$West_{meanday-7}$				0.531									
$West_{maxday-2}$			0.985										
$West_{rangeday-2}$			0.981										
$West_{stdAHCday-8}$					0.427								
$West_{maxday0-3}$												0.970	
$West_{stdday0-3}$												0.879	
$West_{stdAHCday0-8}$					0.328								
$West_{meanday-1to2}$			0.744										
$Pres_{maxAHCday-8}$					0.395								
$Pres_{meanAHCday-7}$					0.386								
$Pres_{maxday-7-8}$				0.328	0.343								
$Rain_{totalday-7}$				0.888									
$Rain_{stdday-7}$				0.811									
$Rain_{totalday0-5}$							0.951						
$Rain_{totalday0-7}$				0.678			0.635						
$Rain_{maxday0-1}$							0.413						
$Rain_{stdday0-1}$							0.444						
$Rain_{meanAHCday0-5}$							0.819						
$Rain_{stdAHCday-6-7}$				0.393									
$Dew_{meanday-1to-14}$	0.947												

Variable	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23
$Temp_{meanday-6-7}$		0.323								
$Temp_{rangeday-6-7}$		0.657								
$WindS_{maxday-7}$							0.953			
$WindS_{maxday0-5}$				0.887						
$WindS_{maxAHCday-7-8}$							0.380			
$WindS_{meanday-1to-14}$							0.318			
$WindS_{diffday0-2}$				0.482						
$WindS_{diffday0-14}$				0.496						
$North_{maxday0-5}$	0.916									
$North_{rangeday0-5}$	0.920									
$South_{maxAHCday-14}$			0.904							
$South_{maxAHCday0-4}$						0.841				
$Humid_{stdday-6-7}$		0.941								
$Pres_{rangeday0}$						0.615				
$Pres_{rangeday-14}$			0.698							
$Pres_{meanday-5-6}$					0.612					
$Pres_{maxday-7-8}$					0.474					
$Pres_{rangeday-6-7}$					0.518					
$WindD_{maxAHCday-4}$								0.980		
$WindD_{meanAHCday-2-3}$										0.988
$WindD_{stdAHCday-1-2}$									0.987	

Table D.4: Component loadings for 23 rotated components, Period 1.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
$Temp_{stdday-7}$						-0.317		
$Temp_{maxday-7-8}$					0.349			
$Temp_{diffday0-7}$		0.654						
$Temp_{diffday0-14}$		0.683						
$Temp_{minday0}$		0.410						
$WindS_{meanday-1}$				0.656				
$WindS_{meanday-2}$				0.765				
$WindS_{rangeday-1}$				0.633				
$WindS_{rangeday-2}$				0.817				
$WindS_{meanday1-2}$				0.336				
$WindS_{meanday-1to-3}$				0.792				
$WindS_{meanday-1to-14}$	0.317			0.439				
$WindS_{diffday0-2}$				-0.581				
$North_{maxday0}$		0.913						
$North_{maxday-8}$					0.938			
$North_{stdday0}$		0.884						
$North_{stdday-8}$					0.916			
$North_{meanday0-3}$		0.417						
$North_{maxday0-5}$								0.897
$North_{maxday0-6}$		0.514					0.394	0.533
$North_{rangeday0-5}$								0.897
$North_{rangeday0-6}$		0.511					0.401	0.525
$North_{maxAHCday-7-8}$					0.448			
$North_{diffday0-4}$		0.728						
$North_{diffday0-14}$		0.829						
$South_{meanday-7}$						0.510		
$South_{stdday-2}$			0.538					
$South_{meanAHCday-7}$						0.468		
$South_{meanday-1to-2}$			0.482					
$East_{meanAHCday0}$		0.396						
$East_{meanAHCday8}$					0.381			
$West_{meanday-1}$			0.406					
$West_{meanday-7}$						0.445		
$West_{maxday-2}$			0.985					
$West_{rangeday-2}$			0.981					
$West_{stdAHCday-8}$					0.413			
$West_{stdAHCday0-8}$					0.326			
$West_{meanday-1to2}$			0.745					
$Pres_{maxAHCday-8}$					0.398			
$Pres_{meanAHCday-7}$					0.418			
$Pres_{maxday-7-8}$					0.379			
$Rain_{totalday-7}$						0.762		
$Rain_{stdday-7}$						0.689		
$Rain_{totalday0-5}$							0.775	
$Rain_{totalday0-7}$						0.668	0.653	
$Rain_{maxday0-1}$							0.377	
$Rain_{stdday0-1}$							0.405	
$Rain_{meanAHCday0-5}$							0.667	
$Rain_{stdAHCday-6-7}$						0.319		
$Dew_{meanday-7}$	0.888							
$Dew_{minday-4}$	0.825							
$Dew_{minday-6}$	0.877							
$Dew_{maxday-5}$	0.894							
$Dew_{maxday-8}$	0.848							
$Dew_{meanday-1to-8}$	0.972							
$Dew_{meanday-1to-14}$	0.945							
Variable	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
$WindS_{meanday-1}$			0.378					
$WindS_{maxday-7}$	0.722							
$WindS_{maxday-8}$	0.847							
$WindS_{rangeday-1}$			0.573					
$WindS_{maxAHCday-8}$	0.701							
$WindS_{maxday0-5}$			0.568					
$WindS_{maxAHCday-7-8}$	0.529							
$WindS_{meanday-1to-14}$	0.448							
$WindS_{diffday0-2}$			0.566					
$WindS_{diffday0-14}$			0.690					
$South_{maxAHCday-14}$					0.885			
$South_{maxday0-5}$		0.957						
$South_{stdday0-5}$		0.915						
$South_{maxAHC0-4}$		0.347						
$West_{maxday0-3}$				0.861				
$West_{stdday0-3}$				0.778				
$Pres_{rangeday-14}$					0.699			
$WindD_{maxAHCday-4}$						0.986		
$WindD_{meanAHCday-2-3}$								0.987
$WindD_{stdAHCday-1-2}$							0.993	

Table D.5: Component loadings for 16 rotated components, Period 1.

Wind Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
$WindS_{meanday-1}$			0.695				
$WindS_{meanday-2}$			0.730				
$WindS_{maxday-7}$				0.743			
$WindS_{maxday-8}$				0.844			
$WindS_{rangeday-1}$			0.683				
$WindS_{rangeday-2}$			0.772				
$WindS_{maxAHCday-8}$				0.697			
$WindS_{meanday1-2}$			0.328				
$WindS_{maxAHCday-7-8}$				0.527			
$WindS_{meanday-1to-3}$			0.791				
$WindS_{meanday-1to-14}$			0.455	0.467			
$WindS_{diffday0-2}$			-0.498				
$North_{maxday0}$	0.699						
$North_{stdday0}$	0.664						
$North_{meanday0-3}$	0.447						
$North_{maxday0-5}$	0.813						
$North_{maxday0-6}$	0.788						
$North_{rangeday0-5}$	0.809						
$North_{rangeday0-6}$	0.783						
$North_{diffday0-4}$	0.498						
$North_{diffday0-14}$	0.573						
$South_{stdday-2}$		0.529					
$South_{meanday-1to-2}$		0.484					
$West_{meanday-1}$		0.414					
$West_{maxday-2}$		0.986					
$West_{rangeday-2}$		0.981					
$West_{maxday0-3}$	0.306						
$West_{stdday0-3}$	0.305						
$West_{meanday-1to2}$		0.753					
$WindD_{maxAHCday-4}$					0.997		
$WindD_{meanAHCday-2-3}$							0.991
$WindD_{stdAHCday-1-2}$						0.991	

Non-wind Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
$Temp_{meanday-6-7}$						0.351								
$Temp_{maxday-7-8}$							0.347							
$Temp_{rangeday-6-7}$						0.671								
$Temp_{diffday0-7}$		0.914												
$Temp_{diffday0-14}$		0.943												
$Temp_{minday0}$		0.857												
$Humid_{stdday-6-7}$						0.960								
$Pres_{rangeday0}$								0.991						
$Pres_{rangeday-14}$											0.986			
$Pres_{meanAHCday-7}$							0.441							
$Pres_{meanday-5-6}$									0.979					
$Pres_{maxday-7-8}$							0.923							
$Pres_{rangeday-6-7}$							0.479						0.849	
$Pres_{stdday0-5}$								0.426						
$Rain_{totalday-7}$			0.933											
$Rain_{stdday-7}$			0.837											
$Rain_{totalday0-5}$				0.985										
$Rain_{totalday0-7}$			0.795	0.523										
$Rain_{maxday0-1}$				0.377										
$Rain_{stdday0-1}$				0.410										
$Rain_{meanAHCday0-5}$				0.851										
$Rain_{stdAHCday-6-7}$			0.389											
$Sun_{totalday-2-3}$												0.951		
$Sun_{totalday0-7}$										0.986				
$Dew_{meanday-7}$	0.921													
$Dew_{minday-4}$	0.750													0.642
$Dew_{minday-6}$	0.885													
$Dew_{maxday-5}$	0.883													
$Dew_{maxday-8}$	0.876													
$Dew_{meanday-1to-8}$	0.966													
$Dew_{meanday-1to-14}$	0.946													
$Dew_{diffday0-8}$					0.914									

Table D.6: Component loadings for wind and non-wind based rotated components, Period 1.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
$Temp_{meanAHCday0}$	0.400				0.464			
$Temp_{stdAHCday0}$						-0.616		
$North_{meanday0}$					0.324			
$North_{maxday-7}$			-0.460					
$North_{rangeday-7}$			-0.459					
$South_{meanday-4}$				0.306				
$South_{maxday-3}$		0.755						
$South_{maxday-5}$				0.687				
$South_{stdday-3}$		0.718						
$South_{meanday-3-4}$		0.364						
$South_{meanday-4-5}$				0.375				
$South_{maxday0-5}$				0.805				
$South_{rangeday0-5}$				0.796				
$South_{meanday-1to-6}$		0.374		0.416				
$South_{diffday0-4}$	-0.383							
$East_{meanday-1to-3}$		-0.487						
$East_{meanAHCday-3}$		-0.440						
$East_{meanday-3}$		-0.569						
$West_{maxday-2}$		0.418					-0.310	
$West_{maxday-3}$		0.753						
$West_{stdday-3}$		0.693						
$West_{maxAHCday-1}$								-0.435
$West_{meanday-2-3}$		0.419						
$West_{meanday-3-4}$		0.383						
$West_{maxday0-1}$					0.350			
$West_{rangeday0-1}$					0.363			
$West_{meanday-1to-3}$		0.640						
$West_{diffday0-3}$		-0.413						
$Humid_{meanAHCday-8}$						-0.678		
$Humid_{meanday-5-6}$			-0.380					
$Humid_{meanday-6-7}$			-0.396					
$Humid_{maxAHCday-2-3}$							-0.330	
$Humid_{minday-1}$					-0.447			0.826
$Humid_{minday-8}$						0.975		
$Pres_{minday-4}$		-0.634	0.591					
$Pres_{minday-5}$		-0.322	0.833					
$Pres_{maxday-6}$			0.844					
$Pres_{maxday0-8}$							-0.309	
$Pres_{meanday-1to-5}$		-0.434	0.562				0.356	
$Pres_{meanday-1to-14}$			0.612					
$Pres_{diffday0-6}$		0.444	-0.443		-0.518			
$Rain_{maxday0}$	-0.427							
$Rain_{stdAHCday0}$	-0.420							
$Rain_{maxday0-1}$	-0.309							
$Rain_{rangeday0-1}$	-0.309							
$Rain_{diffday0-3}$	-0.302							
$Rain_{diffday0-14}$	-0.387							
$Sun_{diffday0-6}$	0.775							
$Sun_{diffday0-14}$	0.817							
$Sun_{totalday0}$	0.850							
$Sun_{totalday-3}$		-0.323						
$Rad_{diffday0-8}$	0.821							0.302
$Rad_{diffday0-14}$	0.835							
$Rad_{totalday0}$	0.857							
$Rad_{totalday-3}$		-0.317						
$Dew_{meanday-2}$							0.326	
$Dew_{maxday0}$							0.314	
$Dew_{maxday-1}$							0.330	
$Dew_{meanday-1to-2}$							0.340	
$WindD_{maxAHCday-3}$								
$WindD_{stdAHCday0}$					0.719			

Variable	PC9	PC10	PC11	PC12	PC13	PC14	PC15
$Temp_{meanAHCday0-2}$		0.432					
$Humid_{maxAHCday-5}$	0.833						
$Humid_{minday0-2}$		0.940					
$Humid_{rangeday0-3}$						0.910	
$Humid_{maxAHCday0-5}$	0.766						
$WindD_{maxAHCday-3}$			0.922				
$WindD_{stdAHCday0}$					0.418		
$WindD_{maxAHCday-4-5}$							0.986
$WindD_{stdAHCday0-3}$					0.908		
$WindD_{stdAHCday0-6}$				0.919			

Table D.7: Component loadings for 15 rotated components, Period 2.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
$Temp_{stdday-3}$									0.419				
$Temp_{meanAHCday0}$	0.393						0.312						
$Temp_{stdAHCday-8}$										-0.624			
$WindS_{maxday0}$			0.334					0.703					
$WindS_{stdday0}$								0.569					
$WindS_{maxday0-5}$								0.773					
$WindS_{rangeday0-5}$								0.773					
$WindS_{maxAHCday0-3}$								0.386					
$WindS_{maxAHCday0-6}$								0.407					
$North_{meanday0}$							0.418						
$North_{maxday-7}$											0.870		
$North_{rangeday-7}$											0.869		
$North_{maxday-2-3}$				0.308									
$North_{stdAHCday-2-3}$				0.301									
$South_{meanday-4}$						0.341							
$South_{maxday-3}$		0.903											
$South_{maxday-5}$						0.851							
$South_{stdday-3}$		0.872											
$South_{meanday-3-4}$		0.475											
$South_{meanday-4-5}$						0.471							
$South_{maxday0-5}$												0.920	
$South_{rangeday0-5}$												0.922	
$South_{meanday-1to-6}$		0.410				0.503							
$South_{meanday-1to-14}$						0.328							
$South_{diffday0-4}$	-0.400	-0.339											
$East_{meanday-1to-3}$		-0.344		-0.424									
$East_{meanAHCday-3}$		-0.576											
$East_{meanday-3}$		-0.636											
$West_{maxday-2}$				0.856									
$West_{maxday-3}$		0.681		0.407									
$West_{stdday-3}$		0.619		0.381									
$West_{meanday-2-3}$				0.447									
$West_{meanday-3-4}$		0.369											
$West_{minday-1-2}$				0.302									
$West_{minday-4-5}$						0.347							
$West_{maxday0-1}$							0.735		0.328				
$West_{rangeday0-1}$							0.738		0.334				
$West_{meanday-1to-3}$		0.386		0.702									
$West_{diffday0-3}$		-0.314		-0.405									
$Humid_{maxAHCday-5}$													0.821
$Humid_{meanAHCday-8}$										-0.696			
$Humid_{maxAHCday0-5}$													0.840
$Humid_{minday-8}$										0.988			
$Pres_{minday-4}$		-0.401			0.833								
$Pres_{minday-5}$					0.815	-0.335							
$Pres_{maxday-6}$					0.629	-0.519					-0.360		
$Pres_{maxday-2-3}$				0.433									
$Pres_{maxday-4-5}$					-0.361								
$Pres_{meanday-1to-5}$				-0.386	0.773								
$Pres_{meanday-1to-14}$					0.654								
$Pres_{diffday0-6}$				0.408	-0.391		-0.571		0.310				
$Rain_{maxday0}$	-0.465												
$Rain_{stdAHCday0}$	-0.458												
$Rain_{maxday0-1}$	-0.343												
$Rain_{maxday0-2}$	-0.333												
$Rain_{maxday0-3}$	-0.319												
$Rain_{maxday0-4}$	-0.301												
$Rain_{maxday0-5}$	-0.309												
$Rain_{rangeday0-1}$	-0.343												
$Rain_{rangeday0-1}$	-0.333												
$Rain_{stdday0-4}$	-0.308												
$Rain_{maxAHCday0-3}$	-0.302												
$Rain_{meanAHCday0-5}$	-0.332												
$Rain_{meanAHCday0-6}$	-0.309												
$Rain_{diffday0-14}$	-0.399												
$Sun_{diffday0-6}$	0.699								-0.442				
$Sun_{diffday0-14}$	0.769								-0.341				
$Sun_{totalday0}$	0.833												
$Sun_{totalday-3}$									0.675				
$Sun_{totalday-5}$						-0.305							
$Rad_{diffday0-8}$	0.753								-0.433				
$Rad_{diffday0-14}$	0.783								-0.380				
$Rad_{totalday0}$	0.838												
$Rad_{totalday-3}$									0.678				
$Dew_{meanday-2}$			0.898										
$Dew_{maxday0}$			0.840										
$Dew_{maxday-1}$			0.907										
$Dew_{maxday-4}$			0.778										
$Dew_{meanday-1to-2}$			0.936										

Table D.8: Component loadings for 26 rotated components, Period 2.

Variable	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26
$Temp_{meanAHCday0}$							0.323						
$Temp_{meanAHCday-2-3}$					0.310								
$Temp_{meanAHCday0-2}$			0.452										
$North_{meanday-5-6}$		0.371											
$North_{meanday0-14}$				-0.309									
$West_{maxAHCday-1}$	-0.457												
$Humid_{meanday-5-6}$		0.950											
$Humid_{meanday-6-7}$								0.870					
$Humid_{minday0-2}$			0.961										
$Humid_{rangeday0-3}$													0.943
$Humid_{maxAHCday-2-3}$					0.966								
$Humid_{minday1}$	0.913												
$Humid_{maxday-14}$				0.979									
$Pres_{maxday0-8}$									0.989				
$WindD_{maxAHCday-3}$						0.920							
$WindD_{stdAHCday0}$							0.944						
$WindD_{maxAHCday-4-5}$											0.989		
$WindD_{stdAHCday0-3}$												0.963	
$WindD_{stdAHCday0-6}$										0.976			

Table D.8: Component loadings for 26 rotated components, Period 2.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
$Temp_{stdday-3}$									0.406				
$Temp_{meanAHCday0}$	0.391						0.308						
$Temp_{stdAHCday-8}$											-0.623		
$WindS_{maxday0}$			0.334					0.704					
$WindS_{stdday0}$								0.569					
$WindS_{maxday0-5}$								0.774					
$WindS_{rangeday0-5}$								0.772					
$WindS_{maxAHCday0-3}$								0.385					
$WindS_{maxAHCday0-6}$								0.407					
$North_{meanday0}$							0.412						
$North_{maxday-7}$										0.755			
$North_{rangeday-7}$										0.755			
$North_{maxday-2-3}$				0.308									
$North_{stdAHCday-2-3}$				0.301									
$South_{meanday-4}$						0.332							
$South_{maxday-3}$		0.908											
$South_{maxday-5}$						0.851							
$South_{stdday-3}$		0.876											
$South_{meanday-3-4}$		0.476											
$South_{meanday-4-5}$						0.461							
$South_{maxday0-5}$												0.911	
$South_{rangeday0-5}$												0.912	
$South_{meanday-1to-6}$		0.412				0.489							
$South_{meanday-1to-14}$						0.317							
$South_{diffday0-4}$	-0.401	-0.338											
$East_{meanday-1to-3}$		-0.341		-0.426									
$East_{meanAHCday-3}$		-0.578											
$East_{meanday-3}$		-0.637											
$West_{maxday-2}$				0.855									
$West_{maxday-3}$		0.679		0.406									
$West_{stdday-3}$		0.616		0.380									
$West_{meanday-2-3}$				0.446									
$West_{meanday-3-4}$		0.367											
$West_{minday-1-2}$				0.301									
$West_{minday-4-5}$						0.330							
$West_{maxday0-1}$							0.742						
$West_{rangeday0-1}$							0.746						
$West_{meanday-1to-3}$		0.386		0.701									
$West_{diffday0-3}$		-0.313		-0.407									
$Humid_{maxAHCday-5}$													0.817
$Humid_{meanAHCday-8}$											-0.695		
$Humid_{meanday-5-6}$													
$Humid_{meanday-6-7}$										0.778			
$Humid_{maxAHCday0-5}$													0.842
$Humid_{minday-8}$											0.985		
$Pres_{minday-4}$		-0.393			0.835								
$Pres_{minday-5}$					0.805	-0.348							
$Pres_{maxday-6}$					0.604	-0.538				-0.379			
$Pres_{maxday-2-3}$				0.433									
$Pres_{maxday-4-5}$					-0.373								
$Pres_{meanday-1to-5}$				-0.389	0.761								
$Pres_{meanday-1to-14}$					0.636								
$Pres_{diffday0-6}$				0.414	-0.403		-0.559						
$Rain_{maxday0}$	-0.465												
$Rain_{stdAHCday0}$	-0.458												
$Rain_{maxday0-1}$	-0.346												
$Rain_{maxday0-2}$	-0.337												
$Rain_{maxday0-3}$	-0.325												
$Rain_{maxday0-4}$	-0.302												
$Rain_{maxday0-5}$	-0.309												
$Rain_{rangeday0-1}$	-0.346												
$Rain_{rangeday0-2}$	-0.337												
$Rain_{stdday0-4}$	-0.309												
$Rain_{maxAHCday0-3}$	-0.308												
$Rain_{meanAHCday0-5}$	-0.331												
$Rain_{meanAHCday0-6}$	-0.308												
$Rain_{diffday0-14}$	-0.398												
$Sun_{diffday0-6}$	0.684								-0.491				
$Sun_{diffday0-14}$	0.755								-0.387				
$Sun_{totalday0}$	0.822												
$Sun_{totalday-3}$									0.637				
$Sun_{totalday-5}$	0.826												
$Rad_{diffday0-8}$	0.737								-0.481				
$Rad_{diffday0-14}$	0.768								-0.426				
$Rad_{totalday0}$									0.636				
$Rad_{totalday-3}$			0.895										

Table D.9: Component loadings for 25 rotated components, Period 2.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
$Dew_{meanday-2}$			0.837										
$Dew_{maxday0}$			0.904										
$Dew_{maxday-1}$			0.774										
$Dew_{maxday-4}$			0.933										
Variable	**PC14**	**PC15**	**PC16**	**PC17**	**PC18**	**PC19**	**PC20**	**PC21**	**PC22**	**PC23**	**PC24**	**PC25**	
$Temp_{meanAHCday0}$							0.324						
$Temp_{meanAHCday-2-3}$					0.309								
$Temp_{meanAHCday0-2}$				0.451									
$North_{meanday-5-6}$		0.362											
$North_{meanday0-14}$			-0.310										
$West_{maxAHCday-1}$	-0.459												
$Humid_{meanday-5-6}$		0.925											
$Humid_{minday0-2}$				0.961									
$Humid_{rangeday0-3}$												0.943	
$Humid_{maxAHCday-2-3}$					0.966								
$Humid_{minday-1}$	0.913												
$Humid_{maxday-14}$			0.978										
$Pres_{maxday0-8}$								0.989					
$Dew_{meanday-1to-2}$						0.922							
$WindD_{stdAHCday0}$							0.944						
$WindD_{maxAHCday-4-5}$										0.989			
$WindD_{stdAHCday0-3}$											0.963		
$WindD_{stdAHCday0-6}$									0.976				

Table D.9: Component loadings for 25 rotated components, Period 2.

Wind Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
$North_{maxday-7}$					0.306		
$North_{rangeday-7}$					0.306		
$South_{maxday-3}$	0.810						
$South_{maxday-5}$		0.632					
$South_{stdday-3}$	0.763						
$South_{meanday-3-4}$	0.390						
$South_{meanday-4-5}$		0.321					
$South_{maxday0-5}$		0.899					
$South_{rangeday0-5}$		0.891					
$South_{meanday-1to-6}$	0.477	0.341					
$South_{diffday0-4}$	-0.419						
$East_{meanday-1to-3}$	-0.504						
$East_{meanAHCday-3}$	-0.530						
$East_{meanday-3}$	-0.614						
$West_{maxday-2}$	0.445						
$West_{maxday-3}$	0.800						
$West_{stdday-3}$	0.725						
$West_{meanday-2-3}$	0.461						
$West_{meanday-3-4}$	0.412						
$West_{meanday-1to-3}$	0.689						
$West_{diffday0-3}$	-0.496						
$WindD_{maxAHCday-3}$				0.938			
$WindD_{stdAHCday0}$			0.937				
$WindD_{maxAHCday-4-5}$						0.984	
$WindD_{stdAHCday0-3}$							0.925
$WindD_{stdAHCday0-6}$					0.956		

Non-wind Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
$Temp_{stdday-3}$					0.530								
$Temp_{meanAHCday0}$	0.303								-0.333				
$Temp_{stdAHCday-8}$						-0.618							
$Temp_{meanAHCday-2-3}$													0.324
$Temp_{meanAHCday0-2}$										0.446			
$Humid_{maxAHCday-5}$												0.923	
$Humid_{meanAHCday-8}$						-0.696							
$Humid_{minday0-2}$										0.970			
$Humid_{maxAHCday-2-3}$													0.990
$Humid_{maxAHCday0-5}$											0.941		
$Humid_{minday-1}$									0.934				
$Humid_{minday-8}$						0.988							
$Pres_{minday-4}$				0.880									
$Pres_{minday-5}$				0.809				0.507					
$Pres_{maxday-6}$				0.387				0.831					
$Pres_{maxday-4-5}$				-0.584									
$Pres_{meanday-1to-5}$	0.456	-0.341		0.653									
$Pres_{meanday-1to-14}$	0.422	-0.348		0.356				0.432					
$Pres_{diffday0-6}$				-0.380									
$Rain_{maxday0}$	-0.424												
$Rain_{stdAHCday0}$	-0.419												
$Rain_{maxday0-1}$	-0.371												
$Rain_{maxday0-2}$	-0.371												
$Rain_{maxday0-3}$	-0.331												
$Rain_{rangeday0-1}$	-0.371												
$Rain_{rangeday0-2}$	-0.371												
$Rain_{maxAHCday0-3}$	-0.318												
$Rain_{diffday0-14}$	-0.318												
$Sun_{totalday0-2}$							0.870						
$Sun_{diffday0-6}$	0.452	0.777											
$Sun_{diffday0-14}$	0.497	0.786											
$Sun_{totalday0}$	0.576	0.731											
$Sun_{totalday-3}$					0.922								
$Sun_{totalday-5}$												0.338	
$Rad_{diffday0-8}$	0.488	0.812											
$Rad_{diffday0-14}$	0.502	0.818											
$Rad_{totalday0}$	0.573	0.759											
$Rad_{totalday-3}$					0.966								
$Rad_{totalday0-2}$							0.910						
$Dew_{meanday-2}$			0.926										
$Dew_{maxday0}$			0.859										
$Dew_{maxday-1}$			0.924										
$Dew_{maxday-4}$			0.791										
$Dew_{meanday-1to-2}$			0.964										

Non-wind Variable	PC14	PC15	PC16	PC17	PC18	PC19
$Humid_{meanday-5-6}$		0.977				
$Humid_{meanday-6-7}$	0.984					
$Humid_{rangeday0-3}$						0.921
$Humid_{maxday-14}$			0.986			
$Pres_{maxday0-8}$				0.990		
$Pres_{diffday0-6}$					0.813	

Table D.10: Component loadings for wind and non-wind based rotated components, Period 2.

Appendix E

Converting Daily Data to Monthly Data

This appendix contains the full component loading details of the monthly components created from the monthly climate variables. Period 1 components are contained in E.1, and Period 2 details in E.2.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
$Dew^{m}_{mean(mean)}$	-0.098	0.043	0.081	-0.036	0.958	-0.117	0.214	-0.005	-0.004
$Dew^{m}_{mean(min)}$	-0.097	0.064	0.075	-0.024	0.971	-0.101	0.146	-0.018	0.008
$Dew^{m}_{mean(max)}$	-0.087	0.015	0.091	-0.030	0.936	-0.121	0.278	0.012	0.012
$Dew^{m}_{mean(std)}$	0.106	-0.177	0.036	0.001	-0.614	0.022	0.291	0.134	0.074
$East^{m}_{mean(mean)}$	0.784	-0.085	-0.422	-0.049	-0.036	0.107	0.004	0.144	0.057
$East^{m}_{mean(min)}$	0.238	-0.109	-0.639	-0.131	-0.035	-0.001	-0.085	0.147	0.039
$East^{m}_{mean(max)}$	0.902	-0.146	-0.184	-0.038	-0.106	0.046	0.020	0.108	0.048
$East^{m}_{mean(std)}$	0.924	-0.100	-0.077	-0.011	-0.101	0.075	0.071	0.060	0.071
$Humid^{m}_{mean(mean)}$	-0.787	0.282	-0.332	-0.126	0.096	0.231	0.130	0.053	0.292
$Humid^{m}_{mean(min)}$	-0.762	0.356	-0.454	-0.018	0.064	0.186	0.085	0.108	0.162
$Humid^{m}_{mean(max)}$	-0.617	0.190	-0.201	-0.262	0.112	0.234	0.205	-0.010	0.584
$Humid^{m}_{mean(std)}$	0.737	-0.349	0.477	-0.090	-0.001	-0.101	0.025	-0.193	0.133
$North^{m}_{mean(mean)}$	0.680	-0.404	0.086	0.059	-0.005	-0.255	-0.069	0.057	-0.188
$North^{m}_{mean(min)}$	0.098	-0.308	-0.131	0.057	-0.022	-0.139	-0.058	0.024	-0.071
$North^{m}_{mean(max)}$	0.663	-0.426	0.257	0.044	0.077	-0.267	-0.052	-0.025	-0.134
$North^{m}_{mean(std)}$	0.695	-0.398	0.302	0.015	0.038	-0.253	-0.030	-0.039	-0.102
$Pres^{m}_{mean(mean)}$	-0.016	-0.209	-0.211	-0.137	-0.113	0.934	-0.061	0.052	0.005
$Pres^{m}_{mean(min)}$	-0.028	-0.191	-0.288	-0.133	-0.137	0.908	-0.081	0.053	0.007
$Pres^{m}_{mean(max)}$	0.010	-0.221	-0.141	-0.128	-0.081	0.943	-0.058	0.053	0.006
$Pres^{m}_{mean(std)}$	0.133	-0.050	0.567	0.055	0.229	-0.170	0.108	-0.017	-0.014
$Rad^{m}_{mean(total)}$	-0.098	0.034	-0.054	-0.098	0.439	-0.048	0.846	-0.188	-0.168
$Rain^{m}_{mean(mean)}$	-0.228	0.836	-0.007	0.155	0.083	-0.149	-0.016	0.072	0.036
$Rain^{m}_{mean(min)}$	-0.233	0.837	-0.008	0.158	0.079	-0.149	-0.012	0.074	0.034
$Rain^{m}_{mean(max)}$	0.030	0.692	-0.019	0.280	0.088	-0.223	-0.087	-0.097	-0.046
$Rain^{m}_{mean(std)}$	-0.034	0.742	0.001	0.251	0.097	-0.207	-0.083	-0.048	-0.014
$South^{m}_{mean(mean)}$	-0.032	0.752	0.291	-0.074	-0.020	0.029	-0.081	0.031	0.044
$South^{m}_{mean(min)}$	-0.300	0.561	-0.023	-0.022	-0.008	0.005	-0.151	0.010	-0.071
$South^{m}_{mean(max)}$	0.223	0.548	0.684	-0.110	-0.112	0.006	0.054	-0.013	0.028
$South^{m}_{mean(std)}$	0.329	0.441	0.689	-0.126	-0.087	0.011	0.061	-0.003	0.054
$Sun^{m}_{mean(total)}$	0.842	-0.206	0.248	0.086	-0.088	-0.035	-0.016	-0.107	-0.045
$Temp^{m}_{mean(mean)}$	0.960	0.055	-0.137	0.129	-0.071	0.095	0.092	-0.078	0.078
$Temp^{m}_{mean(min)}$	0.923	0.162	-0.246	0.151	-0.051	0.110	0.100	-0.032	0.083
$Temp^{m}_{mean(max)}$	0.958	-0.031	-0.040	0.120	-0.113	0.098	0.103	-0.108	0.076
$Temp^{m}_{mean(std)}$	0.324	-0.515	0.566	-0.028	-0.209	-0.029	0.029	-0.226	0.010
$West^{m}_{mean(mean)}$	-0.265	0.441	0.592	0.040	0.085	-0.346	-0.192	0.037	-0.039
$West^{m}_{mean(min)}$	-0.438	0.651	0.183	-0.073	0.035	-0.100	-0.095	0.118	0.072
$West^{m}_{mean(max)}$	0.071	0.169	0.791	0.037	0.102	-0.367	-0.111	-0.051	-0.052
$West^{m}_{mean(std)}$	0.202	0.018	0.768	0.049	0.095	-0.393	-0.098	-0.094	-0.075
$WindS^{m}_{mean(mean)}$	0.037	0.033	0.023	0.074	0.274	-0.021	0.775	-0.036	0.014
$WindS^{m}_{mean(min)}$	-0.025	0.122	-0.070	0.145	0.250	0.096	0.150	0.074	0.050
$WindS^{m}_{mean(max)}$	0.103	-0.094	0.036	0.093	0.042	-0.043	0.826	0.057	0.099
$WindS^{m}_{mean(std)}$	0.110	-0.145	0.043	0.039	-0.010	-0.060	0.805	-0.013	0.099

Table E.1: Principal component loadings for monthly components, Period 1.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
$Dew^m_{min(min)}$	-0.097	0.064	0.075	-0.024	0.971	-0.101	0.146	-0.018	0.008
$East^m_{min(min)}$	0.238	-0.109	-0.639	-0.131	-0.035	-0.001	-0.085	0.147	0.039
$Humid^m_{min(min)}$	-0.782	0.356	-0.454	-0.018	0.064	0.186	0.085	0.108	0.162
$North^m_{min(min)}$	0.098	-0.308	-0.131	0.057	-0.022	-0.139	-0.058	0.024	-0.071
$Pres^m_{min(min)}$	-0.028	-0.191	-0.288	-0.133	-0.137	0.908	-0.081	0.053	0.007
$Rad^m_{min(total)}$	-0.098	0.034	-0.054	-0.098	0.439	-0.048	0.846	-0.188	-0.168
$Rain^m_{min(total)}$	-0.228	0.836	-0.007	0.155	0.083	-0.149	-0.016	0.072	0.036
$South^m_{min(min)}$	-0.300	0.561	-0.023	-0.022	-0.008	0.005	-0.151	0.010	-0.071
$Sun^m_{min(total)}$	0.842	-0.206	0.248	0.086	-0.088	-0.035	-0.016	-0.107	-0.045
$Temp^m_{min(min)}$	0.923	0.162	-0.246	0.151	-0.051	0.110	0.100	-0.032	0.083
$West^m_{min(min)}$	-0.438	0.651	0.183	-0.073	0.035	-0.100	-0.095	0.118	0.072
$WindS^m_{min(min)}$	-0.025	0.122	-0.070	0.145	0.250	0.096	0.150	0.074	0.050
$Dew^m_{max(max)}$	-0.087	0.015	0.091	-0.030	0.936	-0.121	0.278	0.012	0.012
$East^m_{max(max)}$	0.902	-0.146	-0.184	-0.038	-0.106	0.046	0.020	0.108	0.048
$Humid^m_{max(max)}$	-0.617	0.190	-0.201	-0.262	0.112	0.234	0.205	-0.010	0.584
$North^m_{max(max)}$	0.663	-0.426	0.257	0.044	0.077	-0.267	-0.052	-0.025	-0.134
$Pres^m_{max(max)}$	0.010	-0.221	-0.141	-0.128	-0.081	0.943	-0.058	0.053	0.006
$Rad^m_{max(total)}$	-0.098	0.034	-0.054	-0.098	0.439	-0.048	0.846	-0.188	-0.168
$Rain^m_{max(total)}$	-0.228	0.836	-0.007	0.155	0.083	-0.149	-0.016	0.072	0.036
$Rain^m_{max(max)}$	0.030	0.892	-0.019	0.280	0.088	-0.223	-0.087	-0.097	-0.046
$South^m_{max(max)}$	0.223	0.548	0.684	-0.110	-0.112	0.006	0.054	-0.013	0.028
$Sun^m_{max(total)}$	0.842	-0.206	0.248	0.086	-0.088	-0.035	-0.016	-0.107	-0.045
$Temp^m_{max(max)}$	0.958	-0.031	-0.040	0.120	-0.113	0.098	0.103	-0.108	0.076
$West^m_{max(max)}$	0.071	0.169	0.791	0.037	0.102	-0.367	-0.111	-0.051	-0.052
$WindS^m_{max(max)}$	0.103	-0.094	0.036	0.093	0.042	-0.043	0.826	0.057	0.099
$Temp^m_{-diff}$	0.248	0.054	0.025	-0.187	-0.156	-0.023	-0.006	0.071	-0.036
$WindS^m_{-diff}$	-0.126	0.131	0.025	-0.182	0.288	-0.022	0.116	-0.018	-0.245
$North^m_{-diff}$	-0.097	0.070	-0.054	0.185	-0.105	0.021	0.040	0.652	0.024
$South^m_{-diff}$	0.078	-0.423	-0.125	0.109	-0.236	-0.059	-0.131	-0.178	-0.059
$East^m_{-diff}$	0.019	0.165	-0.014	0.619	-0.044	-0.056	-0.046	0.020	-0.027
$West^m_{-diff}$	-0.119	-0.339	-0.365	-0.349	-0.144	-0.017	-0.306	0.155	0.156
$Humid^m_{-diff}$	-0.062	-0.057	-0.073	-0.188	0.088	0.176	0.093	0.898	0.054
$Pres^m_{-diff}$	0.067	0.057	-0.173	-0.837	0.086	-0.047	-0.028	0.016	-0.212
$Rain^m_{-diff}$	0.062	-0.552	-0.037	-0.318	-0.024	0.063	-0.105	-0.284	0.036
Sun^m_{-diff}	0.253	0.110	0.059	0.543	-0.051	-0.057	0.085	-0.087	-0.083
Rad^m_{-diff}	0.145	-0.036	0.103	0.782	-0.073	-0.206	-0.118	0.044	0.027
Dew^m_{-diff}	0.243	-0.086	0.018	-0.062	0.075	0.010	-0.016	-0.044	-0.014
$Temp^m_{+diff}$	-0.066	-0.067	0.006	0.111	-0.103	-0.305	0.065	-0.141	-0.047
$WindS^m_{+diff}$	0.125	0.091	-0.153	0.173	-0.251	-0.011	-0.174	-0.024	0.286
$North^m_{+diff}$	0.027	-0.018	-0.012	0.062	-0.077	-0.106	0.150	-0.597	0.079
$South^m_{+diff}$	-0.003	0.343	0.270	-0.341	0.033	0.256	0.015	0.032	0.112
$East^m_{+diff}$	-0.012	-0.198	-0.216	-0.220	0.108	-0.008	0.146	0.027	0.068
$West^m_{+diff}$	0.038	0.394	0.150	0.407	0.109	-0.025	0.154	-0.053	-0.121
$Humid^m_{+diff}$	-0.059	-0.052	0.115	-0.070	0.153	0.119	0.028	-0.681	-0.053
$Pres^m_{+diff}$	-0.013	0.007	0.154	0.811	0.029	0.006	0.112	-0.046	0.197
$Rain^m_{+diff}$	-0.061	0.437	0.088	0.179	-0.039	-0.075	0.105	0.153	-0.072
Sun^m_{+diff}	-0.154	0.124	0.094	-0.305	-0.012	0.326	0.059	-0.010	0.183
Rad^m_{+diff}	-0.058	-0.027	0.057	-0.710	-0.043	0.201	0.093	-0.135	0.078
Dew^m_{+diff}	0.047	0.127	-0.246	0.018	-0.017	-0.085	-0.009	-0.228	-0.105
$WindD^m_C$	0.003	0.078	-0.037	0.838	-0.129	-0.049	-0.097	0.279	-0.019
$WindD^m_N$	0.030	-0.108	0.069	-0.087	-0.057	-0.006	0.072	-0.482	0.161
$WindD^m_{NE}$	0.013	0.017	-0.354	-0.302	0.167	-0.333	-0.345	0.029	0.203
$WindD^m_E$	-0.038	-0.156	0.139	-0.561	-0.118	0.336	0.180	-0.050	-0.062
$WindD^m_S$	-0.120	-0.049	0.015	-0.507	0.026	0.101	-0.210	0.135	0.041
$WindD^m_{SW}$	-0.050	0.266	-0.043	0.596	0.140	0.023	0.285	0.141	0.020
$WindD^m_W$	0.056	0.189	0.034	0.499	0.040	-0.026	0.090	0.177	-0.143
$WindD^m_{NW}$	0.129	-0.138	0.317	-0.109	0.017	-0.030	0.068	-0.516	-0.158
$WindD^m_M$	0.087	-0.162	0.032	-0.186	0.049	-0.255	0.021	-0.389	-0.030

Table E.1: Principal component loadings for monthly components, Period 1.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
$Dew^m_{mean(mean)}$	-0.014	0.949	0.056	-0.026	0.018	0.022	-0.038	-0.199	-0.046	0.071	0.060
$Dew^m_{mean(min)}$	-0.045	0.932	0.071	-0.023	0.031	0.012	-0.043	-0.185	-0.041	0.084	0.062
$Dew^m_{mean(max)}$	0.006	0.955	0.042	-0.023	0.003	0.037	-0.024	-0.207	-0.056	0.042	0.077
$Dew^m_{mean(std)}$	0.166	-0.330	-0.109	-0.005	-0.098	0.053	0.086	-0.115	-0.011	-0.164	0.066
$East^m_{mean(mean)}$	0.859	-0.004	-0.070	0.228	0.010	0.082	0.003	0.003	-0.170	-0.018	-0.078
$East^m_{mean(min)}$	0.413	-0.106	0.283	0.246	0.025	0.011	-0.074	0.293	-0.131	0.026	-0.065
$East^m_{mean(max)}$	0.866	0.029	-0.242	0.075	0.007	0.087	0.019	-0.143	-0.153	-0.042	-0.088
$East^m_{mean(std)}$	0.842	0.022	-0.269	0.052	0.007	0.108	0.024	-0.201	-0.145	-0.054	-0.102
$Humid^m_{mean(mean)}$	-0.528	-0.021	0.666	0.461	0.039	-0.079	-0.059	0.020	0.166	-0.035	-0.016
$Humid^m_{mean(min)}$	-0.485	-0.100	0.758	0.378	0.018	-0.018	-0.029	0.007	0.164	-0.032	-0.036
$Humid^m_{mean(max)}$	-0.347	0.079	0.465	0.484	0.121	-0.076	-0.073	-0.036	0.119	-0.139	0.090
$Humid^m_{mean(std)}$	0.518	0.133	-0.723	-0.230	0.019	-0.001	0.002	0.023	-0.156	0.005	0.078
$North^m_{mean(mean)}$	0.591	-0.092	-0.472	-0.351	-0.037	0.023	0.040	-0.044	-0.024	-0.016	-0.067
$North^m_{mean(min)}$	0.518	-0.064	-0.023	0.041	-0.025	-0.110	-0.011	0.206	0.089	0.135	-0.134
$North^m_{mean(max)}$	0.532	0.009	-0.551	-0.419	-0.027	0.037	0.038	-0.141	-0.031	-0.058	-0.032
$North^m_{mean(std)}$	0.521	0.024	-0.557	-0.403	-0.020	0.064	0.035	-0.191	-0.040	-0.082	-0.003
$Pres^m_{mean(mean)}$	-0.053	-0.023	-0.106	0.969	0.013	-0.108	-0.043	-0.109	0.136	-0.006	-0.002
$Pres^m_{mean(min)}$	-0.037	-0.032	-0.066	0.970	0.015	-0.128	-0.046	-0.108	0.125	0.005	0.032
$Pres^m_{mean(max)}$	-0.051	-0.025	-0.144	0.960	0.007	-0.077	-0.037	-0.114	0.147	-0.012	-0.022
$Pres^m_{mean(std)}$	-0.050	0.051	-0.295	-0.496	-0.036	0.285	0.042	0.035	0.043	-0.077	-0.240
$Rad^m_{mean(total)}$	-0.012	0.895	0.026	0.008	-0.057	-0.095	0.011	0.344	-0.086	-0.105	0.099
$Rain^m_{mean(mean)}$	-0.175	0.066	0.667	-0.095	-0.059	-0.041	0.056	-0.058	0.019	-0.038	-0.023
$Rain^m_{mean(min)}$	-0.190	0.065	0.672	-0.077	-0.056	-0.043	0.053	-0.054	0.011	-0.027	-0.023
$Rain^m_{mean(max)}$	0.007	0.082	0.606	-0.128	-0.095	-0.069	0.077	-0.037	-0.039	-0.083	0.100
$Rain^m_{mean(std)}$	-0.019	0.098	0.621	-0.133	-0.097	-0.067	0.079	-0.026	-0.033	-0.084	0.076
$South^m_{mean(mean)}$	0.091	0.008	0.380	-0.036	-0.008	0.350	-0.005	-0.151	0.079	0.012	-0.027
$South^m_{mean(min)}$	-0.159	-0.032	0.500	0.022	-0.039	0.004	0.007	-0.152	0.019	0.054	-0.172
$South^m_{mean(max)}$	0.127	0.005	0.023	-0.271	0.030	0.306	0.013	-0.287	0.105	-0.093	-0.003
$South^m_{mean(std)}$	0.234	0.004	-0.019	-0.228	0.044	0.319	-0.013	-0.230	0.092	-0.077	0.034
$Sun^m_{mean(total)}$	0.661	-0.194	-0.496	-0.107	0.040	0.075	-0.030	-0.069	-0.073	-0.017	0.000
$Temp^m_{mean(mean)}$	0.932	-0.019	-0.264	0.034	-0.003	-0.189	-0.002	-0.040	0.016	0.059	0.033
$Temp^m_{mean(min)}$	0.947	0.001	-0.135	0.081	-0.011	-0.209	-0.001	-0.018	0.011	0.070	0.061
$Temp^m_{mean(max)}$	0.890	-0.047	-0.371	0.000	0.001	-0.186	-0.008	-0.052	0.033	0.066	0.019
$Temp^m_{mean(std)}$	0.061	-0.185	-0.772	-0.204	0.037	0.036	-0.020	-0.123	0.078	0.034	-0.078
$West^m_{mean(mean)}$	-0.372	-0.025	0.106	-0.627	-0.080	0.209	0.104	-0.146	0.018	-0.022	-0.055
$West^m_{mean(min)}$	-0.480	0.011	0.566	0.028	-0.032	0.081	0.022	-0.066	0.051	-0.123	-0.044
$West^m_{mean(max)}$	-0.110	0.026	-0.252	-0.766	-0.062	0.188	0.076	-0.241	0.020	-0.058	-0.030
$West^m_{mean(std)}$	0.017	-0.001	-0.309	-0.752	-0.058	0.192	0.070	-0.237	0.010	-0.039	-0.031
$WindS^m_{mean(mean)}$	-0.023	0.717	0.031	0.014	-0.074	0.053	0.026	0.364	0.003	-0.197	0.008
$WindS^m_{mean(min)}$	-0.175	0.199	-0.027	-0.109	0.012	0.114	0.026	0.323	-0.066	0.047	0.029
$WindS^m_{mean(max)}$	0.069	0.857	0.104	0.029	-0.105	0.044	0.051	0.325	0.043	-0.299	-0.064
$WindS^m_{mean(std)}$	0.098	0.858	0.122	0.021	-0.104	-0.036	0.037	0.269	0.044	-0.319	-0.113
$Dew^m_{min(min)}$	-0.116	0.881	0.013	-0.049	-0.032	-0.056	-0.012	0.142	-0.079	0.012	0.046
$East^m_{min(min)}$	0.542	-0.167	0.096	-0.107	0.022	0.343	0.005	0.033	0.010	0.098	0.138
$Humid^m_{min(min)}$	-0.333	-0.128	0.202	0.234	0.042	-0.171	-0.095	0.019	0.820	0.023	0.250
$North^m_{min(min)}$	0.340	0.024	0.063	-0.164	0.075	0.043	-0.176	0.038	-0.066	0.280	-0.071
$Pres^m_{min(min)}$	-0.065	0.085	-0.060	0.253	-0.020	-0.122	-0.041	-0.072	-0.031	0.009	0.938
$Rad^m_{min(total)}$	-0.029	0.841	-0.030	0.088	-0.029	-0.088	-0.030	0.724	-0.121	-0.030	0.028
$Rain^m_{min(total)}$	-0.016	-0.192	0.070	-0.080	0.014	-0.217	-0.086	-0.046	0.506	0.042	-0.091
$South^m_{min(min)}$	-0.027	-0.169	0.007	0.115	-0.020	-0.015	-0.016	-0.153	0.309	0.072	-0.051
$Sun^m_{min(total)}$	0.472	-0.151	-0.046	-0.275	0.007	0.310	0.039	-0.055	-0.645	-0.009	-0.075
$Temp^m_{min(min)}$	0.915	0.012	-0.105	-0.078	0.013	0.209	-0.052	-0.002	-0.085	0.069	0.151
$West^m_{min(min)}$	-0.272	-0.077	-0.094	0.096	0.024	-0.167	-0.008	-0.160	0.435	-0.086	-0.227
$WindS^m_{min(min)}$	0.090	0.209	-0.047	-0.128	0.004	0.111	-0.055	0.338	-0.001	0.180	-0.031
$Dew^m_{max(max)}$	-0.021	0.848	0.017	-0.034	-0.024	0.033	-0.010	0.295	-0.071	0.007	-0.050
$East^m_{max(max)}$	0.160	-0.091	-0.018	-0.187	0.010	0.816	0.005	-0.067	-0.428	-0.023	0.021
$Humid^m_{max(max)}$	0.027	-0.035	0.045	0.246	0.040	-0.860	-0.030	-0.071	0.105	0.335	0.178
$North^m_{max(max)}$	-0.115	0.062	0.018	-0.196	-0.026	0.810	0.002	0.013	-0.236	-0.176	0.184
$Pres^m_{max(max)}$	0.292	-0.008	-0.138	-0.031	0.012	-0.161	-0.093	-0.034	0.201	0.851	0.013
$Rad^m_{max(total)}$	-0.070	0.518	-0.070	0.099	-0.032	-0.042	-0.019	0.824	-0.087	0.045	0.075
$Rain^m_{max(total)}$	0.123	-0.023	0.159	-0.009	0.002	-0.526	0.006	-0.035	-0.013	-0.115	0.058
$Rain^m_{max(max)}$	0.226	-0.005	0.102	-0.012	0.029	-0.366	-0.018	-0.027	-0.079	-0.172	0.076
$South^m_{max(max)}$	0.281	0.125	0.225	-0.184	0.065	-0.056	0.147	-0.177	-0.004	-0.328	-0.059
$Sun^m_{max(total)}$	0.150	-0.116	0.035	-0.342	0.007	0.689	0.043	0.045	-0.311	-0.050	-0.111
$Temp^m_{max(max)}$	0.492	0.005	-0.092	-0.220	0.023	0.685	-0.041	0.073	-0.417	-0.016	0.104
$West^m_{max(max)}$	-0.215	0.341	0.111	-0.181	-0.032	0.074	0.135	-0.153	0.141	-0.560	-0.100
$WindS^m_{max(max)}$	0.020	0.215	-0.122	0.042	-0.042	0.062	0.009	0.687	-0.086	-0.212	-0.050

Table E.2: Principal component loadings for monthly components, Period 2.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
$Temp^m_{-diff}$	-0.101	-0.037	0.029	0.021	0.193	-0.011	0.307	-0.002	0.053	-0.053	0.024
$WindS^m_{-diff}$	-0.007	0.066	0.019	-0.081	-0.050	-0.021	0.101	-0.068	-0.016	-0.065	0.028
$North^m_{-diff}$	0.017	-0.082	-0.002	-0.013	0.264	0.068	0.563	-0.068	-0.034	-0.081	0.094
$South^m_{-diff}$	0.083	-0.018	-0.019	-0.104	0.045	0.009	-0.128	-0.133	0.125	0.016	0.038
$East^m_{-diff}$	-0.001	-0.044	-0.037	-0.018	0.659	0.001	0.175	-0.051	-0.028	-0.025	0.012
$West^m_{-diff}$	-0.014	0.033	0.032	-0.090	-0.317	-0.005	0.178	-0.052	0.114	-0.007	0.082
$Humid^m_{-diff}$	0.097	-0.031	0.051	0.020	0.054	0.021	0.943	-0.005	0.003	0.077	-0.031
$Pres^m_{-diff}$	0.071	0.024	0.021	-0.180	-0.608	0.131	0.170	0.050	0.051	-0.176	0.034
$Rain^m_{-diff}$	-0.025	-0.132	0.015	-0.018	-0.033	0.016	-0.206	0.026	0.037	0.015	0.106
Sun^m_{-diff}	0.030	0.017	0.024	0.026	0.655	-0.014	0.155	0.054	-0.024	0.013	-0.002
Rad^m_{-diff}	0.008	0.009	0.024	0.059	0.876	0.006	0.189	0.046	-0.025	-0.009	0.043
Dew^m_{-diff}	0.065	-0.097	-0.030	-0.018	0.078	-0.030	-0.072	0.064	0.061	0.021	0.010
$Temp^m_{+diff}$	-0.057	-0.061	0.129	0.129	-0.091	0.024	-0.429	0.091	-0.074	0.030	0.076
$WindS^m_{+diff}$	0.033	-0.033	-0.062	0.024	0.011	0.022	-0.147	0.081	0.040	0.153	-0.111
$North^m_{+diff}$	0.019	0.041	-0.041	0.025	0.007	-0.018	-0.572	-0.009	0.056	0.053	-0.004
$South^m_{+diff}$	-0.024	-0.030	0.000	0.084	-0.234	-0.078	-0.003	0.075	-0.085	0.039	-0.084
$East^m_{+diff}$	0.003	0.015	0.043	0.109	-0.433	-0.057	-0.221	0.035	0.044	0.085	-0.078
$West^m_{+diff}$	-0.045	-0.065	-0.119	0.104	0.197	-0.008	-0.081	0.139	-0.066	-0.041	0.000
$Humid^m_{+diff}$	-0.010	0.102	-0.021	0.056	-0.315	-0.014	-0.633	-0.133	0.080	0.020	-0.043
$Pres^m_{+diff}$	0.021	0.000	-0.126	0.090	0.568	-0.003	-0.171	-0.050	0.085	0.065	-0.030
$Rain^m_{+diff}$	0.044	0.079	0.067	-0.087	-0.077	-0.033	0.223	-0.002	-0.033	0.011	-0.034
Sun^m_{+diff}	-0.037	-0.023	-0.073	0.012	-0.409	0.057	-0.005	-0.029	-0.093	-0.014	-0.036
Rad^m_{+diff}	-0.031	0.003	0.033	-0.011	-0.789	0.046	-0.208	-0.055	-0.049	0.010	-0.069
Dew^m_{+diff}	0.001	0.087	0.007	0.138	-0.029	0.089	-0.148	-0.094	0.068	-0.007	0.056
$WindD^m_O$	0.019	0.087	0.020	-0.060	0.673	0.053	0.253	-0.017	-0.020	-0.042	0.014
$WindD^m_N$	0.045	0.111	0.003	0.022	0.022	-0.023	-0.318	-0.058	0.006	0.037	-0.046
$WindD^m_{NE}$	-0.007	-0.017	-0.015	0.020	0.028	-0.046	0.058	0.214	0.029	0.096	-0.020
$WindD^m_E$	0.010	-0.032	-0.002	0.019	-0.644	0.005	-0.149	-0.118	0.003	-0.043	0.051
$WindD^m_S$	0.008	0.038	0.069	-0.019	-0.578	-0.021	-0.040	0.021	0.000	-0.011	0.082
$WindD^m_{SW}$	-0.042	0.004	0.027	-0.041	0.296	0.010	0.404	0.107	-0.050	-0.026	-0.115
$WindD^m_W$	-0.032	-0.162	-0.099	0.083	0.511	-0.007	-0.005	-0.132	-0.002	0.068	-0.048
$WindD^m_{NW}$	0.036	-0.073	-0.061	-0.047	-0.147	-0.039	-0.596	-0.023	-0.040	-0.023	-0.034
$WindD^m_M$	-0.033	0.062	0.045	0.086	-0.131	0.005	0.016	0.061	0.110	0.044	0.069

Table E.2: Principal component loadings for monthly components, Period 2.

Appendix F

Poisson Regression

This appendix contains detailed information for various Poisson regression models discussed in Chapter 6.

Variables	$\hat{\beta}$ (se)	D	AIC	SC	P
$sin(\frac{2\pi t}{12})$	-0.318 (0.066)	229.217	61.372	95.917	196.494
$cos(\frac{2\pi t}{12})$	-0.447 (0.067)				
NAR	-0.0002 (0.000)				
$Humid^{m}_{mean(max)}$	0.017 (0.005)				
$sin(\frac{2\pi t}{12})$	-0.318 (0.066)	230.121	62.276	104.821	197.483
$cos(\frac{2\pi t}{12})$	-0.448 (0.067)				
NAR	-0.0002 (0.000)				
$Pres^{m}_{mean(mean)}$	0.002 (0.000)				
$sin(\frac{2\pi t}{12})$	-0.318 (0.066)	230.138	62.293	96.838	197.490
$cos(\frac{2\pi t}{12})$	-0.448 (0.067)				
NAR	-0.0002 (0.000)				
$Pres^{m}_{mean(min)}$	0.002 (0.001)				
$sin(\frac{2\pi t}{12})$	-0.318 (0.066)	230.092	62.248	104.793	197.442
$cos(\frac{2\pi t}{12})$	-0.448 (0.067)				
NAR	-0.0002 (0.000)				
$Pres^{m}_{mean(max)}$	0.002 (0.000)				
$sin(\frac{2\pi t}{12})$	-0.334 (0.067)	238.771	70.926	113.471	218.629
$cos(\frac{2\pi t}{12})$	-0.446 (0.068)				
NAR	0.0001 (0.000)				
$West^{m}_{mean(mean)}$	0.119 (0.061)				
$sin(\frac{2\pi t}{12})$	-0.317 (0.066)	229.960	62.115	104.660	197.068
$cos(\frac{2\pi t}{12})$	-0.448 (0.067)				
NAR	-0.0002 (0.000)				
$Pres^{m}_{min(min)}$	0.002 (0.000)				
$sin(\frac{2\pi t}{12})$	-0.334 (0.067)	238.954	71.109	113.654	218.800
$cos(\frac{2\pi t}{12})$	-0.448 (0.068)				
NAR	0.0001 (0.000)				
$Dew^{m}_{max(max)}$	0.035 (0.018)				
$sin(\frac{2\pi t}{12})$	-0.317 (0.066)	229.761	61.916	104.461	196.782
$cos(\frac{2\pi t}{12})$	-0.448 (0.067)				
NAR	-0.0002 (0.000)				
$Pres^{m}_{max(max)}$	0.002 (0.000)				
$sin(\frac{2\pi t}{12})$	-0.333 (0.067)	238.588	70.743	113.288	216.034
$cos(\frac{2\pi t}{12})$	-0.445 (0.067)				
NAR	0.0001 (0.000)				
$West^{m}_{max(max)}$	0.049 (0.024)				
$sin(\frac{2\pi t}{12})$	-0.326 (0.067)	234.768	66.923	109.468	208.686
$cos(\frac{2\pi t}{12})$	-0.436 (0.067)				
NAR	0.0001 (0.003)				
$WindS^{m}_{-diff}$	-0.161 (0.057)				
$sin(\frac{2\pi t}{12})$	-0.323 (0.067)	237.059	69.214	111.759	212.835
$cos(\frac{2\pi t}{12})$	-0.456 (0.067)				
NAR	0.0001 (0.000)				
$WindS^{m}_{+diff}$	0.102 (0.043)				
$sin(\frac{2\pi t}{12})$	-0.323 (0.067)	238.203	217.073	70.358	112.903
$cos(\frac{2\pi t}{12})$	-0.459 (0.068)				
NAR	0.0001 (0.000)				
$North^{m}_{+diff}$	0.124 (0.059)				
$sin(\frac{2\pi t}{12})$	-0.329 (0.067)	239.234	218.857	71.388	113.933
$cos(\frac{2\pi t}{12})$	-0.454 (0.068)				
NAR	0.0001 (0.000)				
Dew^{m}_{+diff}	0.1731 (0.094)				
$sin(\frac{2\pi t}{12})$	-0.341 (0.067)	237.831	221.897	69.986	104.531
$cos(\frac{2\pi t}{12})$	-0.444 (0.067)				
NAR	0.0002 (0.000)				
$WindD^{md}_{NW}$	1.647 (0.741)				

Table F.1: Detailed model descriptors for various univariate climate models with baseline form given by P2(4), Period 2.

Appendix G

Models with Intercept and NAR

This appendix is included for completeness, and presents analogous models to those in the main body of the thesis, where those in the main body do not contain both an intercept and NAR (number of infants at risk). The models in this section all contain an intercept term and NAR. It must be noted that significance of the regressor in these models is not examined.

Model	Variable	$\hat{\beta}$ (se)	D	AIC	SC	P
Model 4	Intercept	1.284 (0.501)	711.584	723.584	743.493	197.011
	$sin(\frac{2\pi t}{12})$	-0.310 (0.067)				
	$cos(\frac{2\pi t}{12})$	-0.488 (0.072)				
	NAR	-0.0002 (0.0001)				
	$West^{m}_{mean(mean)}$	0.051 (0.066)				
	$WindD^{md}_{NW}$	1.224 (0.757)				
Model 5	Intercept	1.275 (0.487)	710.813	722.813	742.722	195.955
	$sin(\frac{2\pi t}{12})$	-0.309 (0.067)				
	$cos(\frac{2\pi t}{12})$	-0.491 (0.072)				
	NAR	-0.0002 (0.0001)				
	$Dew^{m}_{max(max)}$	0.022 (0.019)				
	$WindD^{md}_{NW}$	1.266 (0.763)				
Model 6	Intercept	1.275 (0.517)	711.737	723.737	743.646	196.794
$sin(\frac{2\pi t}{12})$	-0.310 (0.067)					
	$cos(\frac{2\pi t}{12})$	-0.489 (0.072)				
	NAR	-0.0002 (0.0001)				
	$West^{m}_{max(max)}$	0.018 (0.027)				
	$WindD^{md}_{NW}$	1.235 (0.763)				

Table G.1: Details for various multivariate climate models, containing intercept and NAR, Period 2 ('d' corresponds to deseasoned variables). These models are analogous to those in Table 6.4, page 128.

		NB	ZIP	EPP[j]
$\log(\lambda(\mathbf{x}^m))$	Intercept	10.356	10.331	16.327
		(8.181)	(8.172)	(12.560)
	$Temp^{m}_{meanMA30}$	-0.058	-0.058	-0.050
		(0.030)	(0.030)	(0.045)
	NAR	-0.002	-0.002	-0.001
		(0.001)	(0.001)	(0.002)
$\text{logit}(p(\mathbf{z}^m))$	$Temp^{m}_{meanMA30}$	$d = 0.0014$	-5.269	$b = 2.725$
			(136295.7)	$c = -0.491$
m		4	4	5
D		161.643	164.881	164.469
AIC		169.643	172.881	174.469
SC		182.916	186.153	191.059
P		46.554	42.010	43.809

Table G.2: Baseline model parameter estimates (standard errors) for the mixture methods, including intercept and NAR in the models, Period 1. These models are analogous to those in Table 7.3, page 150.

[j] Modelling $\log(a) = \mathbf{X}^m\beta$.

		NB	ZIP	Hurdle	EPP[i]
$\log(\lambda(\mathbf{x}^m))$	Intercept	8.919	10.254	25.957	9.808
		(8.236)	(0.902)	(4.518)	(27.978)
	$Temp^m_{meanMA30}{}^j$	-0.053	-0.047	-0.071	-0.068
		(0.031)	(0.031)	(0.002)	(0.106)
	NAR^t	-0.002	-0.002	-0.005	-0.002
		(0.001)	(0.002)	(0.004)	(0.005)
	$Dew^m_{+diff}{}^k$	0.287	0.412	0.736	
		(0.269)	(0.260)	(0.013)	
	$Humid^{dm}_{-diff}{}^r$				-0.081
					(0.143)
	$West^{dm}_{+diff}{}^s$				-0.294
					(0.507)
$\text{logit}(p(\mathbf{z}^m))$	Intercept			-1.337	
				(0.333)	
	$Dew^m_{mean(min)}{}^v$	$d = 0.0022$	-0.352		$b = 0.020$
			(0.197)		$c = -0.119$
	$South^{dm}_{-diff}{}^w$			-1.208	
				(0.706)	
m		7	5	6	7
D		154.898	188.543	153.811	152.375
AIC		168.898	198.543	165.811	166.375
SC		192.124	215.134	185.719	190.602
P		41.633	44.480	46.671	40.781

Table G.3: Climate model parameter estimates (standard errors) for the mixture methods, including intercept and NAR in the models, Period 1 ('d' = deseasoned, m corresponds to the number of parameters in the model). These models are analogous to those in Table 7.4, page 152.

[i] Modelling $\log(a) = \mathbf{X}^m\beta$.

[j] $Temp^m_{meanMA30}$ = monthly seasonality measure for Period 1.

[t] NAR = number of infants at risk in each month.

[k] Dew^m_{+diff} = monthly average when dewpoint was higher on $day0$ than it had been over the past week.

[r] $Humid^{dm}_{-diff}$ = monthly average when humidity was lower on $day0$ than it had been over the past week (deseasoned).

[s] $West^{dm}_{+diff}$ = monthly average when the western component of wind velocity was higher on $day0$ than it had been over the past week (deseasoned).

[v] $Dew^m_{mean(min)}$ = monthly average of the daily minimum dewpoint.

[w] $South^{dm}_{-diff}$ = monthly average when the southern component of wind velocity was lower on $day0$ than it had been over the past week (deseasoned).

		NB	EPP[j]
$\log(\lambda(\mathbf{x}^m))$	Intercept	1.619	1.619
		(0.459)	(0.478)
	$\sin(\frac{2\pi t}{12})$	-0.318	-0.317
		(0.067)	(0.069)
	$\cos(\frac{2\pi t}{12})$	-0.448	-0.448
		(0.067)	(0.070)
	NAR	-0.0002	-0.0002
		(0.0001)	(0.0001)
		$d = 0.0065$	$b = 0.333$
			$c = -0.001$
m		5	6
D		714.606	714.605
AIC		724.606	726.605
SC		741.120	746.514
P		193.602	220.006

Table G.4: Baseline model parameter estimates (standard errors) for the mixture methods, including intercept and NAR in the NB and EPP models, Period 2. These models are analogous to those in Table 7.9, page 161.

[j] Modelling $\log(a) = \mathbf{X}^m\beta$.

		NB
$\log(\lambda(\mathbf{x}^m))$	Intercept	1.053
		(0.531)
	$\sin(\frac{2\pi t}{12})$	-0.317
		(0.067)
	$\cos(\frac{2\pi t}{12})$	-0.449
		(0.068)
	NAR[r]	-0.0002
		(0.0001)
	$WindS^m_{mean(std)}$[s]	0.158
		(0.073)
		$d = 0.010$
m		6
D		709.818
AIC		721.818
SC		741.727
P		186.263

Table G.5: Climate model parameter estimates (standard errors) for the NB mixture methods, with intercept and NAR, Period 2. These models are analogous to those in Table 7.10, page 163.

[r] NAR = number of infants at risk of SIDS.

[s] $WindS^m_{mean(std)}$ = monthly average of the daily standard deviation of wind speed.

Model	Parameter	Zeger's method $\hat{\beta}_Z$	SE_Z	Asym. SE_A
Model 2	Intercept	1.300	0.506	0.386
	$\sin(\frac{2\pi t}{12})$	-0.390	0.068	0.067
	$\cos(\frac{2\pi t}{12})$	-0.445	0.067	0.068
	NAR	-0.0001	0.0001	0.0001
	$West^{m}_{mean(mean)}$	0.075	0.066	0.066
	$WindD^{md}_{NW}$	0.749	0.802	0.752
Model 3	Intercept	1.164	1.564	0.416
	$\sin(\frac{2\pi t}{12})$	-0.500	0.201	0.067
	$\cos(\frac{2\pi t}{12})$	-0.472	0.200	0.068
	NAR	-0.0002	0.0003	0.0001
	$Dew^{m}_{max(max)}$	0.049	0.065	0.186
	$WindD^{md}_{NW}$	0.489	2.644	0.758
Model 4	Intercept	1.343	1.492	0.431
	$\sin(\frac{2\pi t}{12})$	-0.458	0.176	0.067
	$\cos(\frac{2\pi t}{12})$	-0.450	0.176	0.068
	NAR	-0.0001	0.0003	0.0001
	$West^{m}_{max(max)}$	0.007	0.042	0.027
	$WindD^{md}_{NW}$	0.279	2.437	0.757

Table G.6: Regression estimates, with corresponding standard errors for the parameter-driven models, containing intercept and NAR. These models are analogous to those in Table 8.1, page 183.

References

A. Zeevi and R. Meir and R. Adler (1998). Nonlinear Models for Time Series using Mixtures of Autoregressive Models. *http://wwww-isl.stanford.edu/ azeevi.*

Abramson, H. (1944). Accidental mechanical suffocation in infants. *Journal of Pediatrics*, 25, 404–413.

Adams, E., Chavez, G., Steen, D., Shah, R., Iyasu, S., & Krous, H. (1998). Changes in the Epidemiologic Profile of Sudden Infant Death Syndrome as Rates Decline Among California Infants: 1990 - 1995. *Pediatrics*, 102(6), 1445–1451.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis.* Wiley Series in Probability and Statistics. New York: John Wiley & Sons, Inc.

Aickin, M. & Gensler, H. (1996). Adjusting for Multiple Testing When Reporting Research Results: The Bonferroni vx Holm Methods. *American Journal of Public Health*, 86(5), 726–728.

Aitkin, M. (1999). A General Maximum Likelihood Analysis of Variance Components in Generalized Linear Models. *Biometrics*, 55, 117–128.

Aitkin, M. & Alfo, M. (2003). Longitudinal analysis of repeated binary data using autoregressive and random effect modelling. *Statistical Modelling*, 3, 291–303.

Alm, B., Norvenius, S., Wennergren, F., Skjaerven, R., Oyen, N., Milerad, J., Wennborg, M., Kjaerbeck, J., Helweg-Larsen, K., Irgens, L., & the Nordic Epidemiological SIDS Study (2001). Changes in the epidemiology of sudden infant death syndrome in Sweden 1973-1996. *Archives of Disease in Childhood*, 84, 24–30.

Alm, B., Wennergren, F., Norvenius, G., Oyen, N., Wennergren, F., Skjaerven, R., Oyen, N., Helweg-Larsen, K., Lagercrantz, H., Irgens, L., & the Nordic Epidemiological SIDS Study (1999). Caffeine and alcohol as risk factors for sudden infant death syndrome. *Archives of Disease in Childhood*, 81, 107–111.

Anderson, H. & Cook, D. (1997). Passive smoking and sudden infant death syndrome: review of the epidemiological evidence. *Thorax*, 52, 1003–1009.

Ansari, T., Sibbons, P., Parsons, A., & Rossi, M. (2002). Quantitative neuropathological analysis of sudden infant death syndrome. *Child: Care, Health and Development*, 28(Supplement), 3–6.

Arneil, G., Brooke, H., Gibson, A., Harvie, A., McIntosh, H., & Patrick, W. (1985). National post-perinatal infant mortality and cot death study, Scotland 1981-82. *Lancet*, 1(8431), 740–743.

Arnestad, M., Andersen, M., & Rognum, T. (1997). Is the use of dummy or carry-cot of importance for sudden infant death? *European Journal of Pediatrics*, 156, 968–970.

Auliciems, A. & Barnes, A. (1987). Sudden infant deaths and clear weather in a subtropical environment. *Social Science in Medicine*, 24(1), 51–56.

Barry, S. & Welsh, A. (2002). Generalized additive modelling and zero inflated count data. *Ecological Modelling*, 157(2-3), 179–188.

Batschelet, E. (1981). *Circular Statistics in Biology*. London: Academic Press.

Beckwith, J. (1970). Discussion of terminology and definition of sudden infant death syndrome. In A. Bergman, J. Beckwith, & C. Ray (Eds.), *Proceedings of the Second International Conference on Causes of Sudden Death in Infants* Seattle: University of Washington Press.

Beckwith, J. (1973). The sudden infant death syndrome. *Current Problems in Pediatrics*, 3(8), 1–36.

Bender, R. & Lange, S. (2001). Adjusting for multiple testing – when and how? *Journal of Clinical Epidemiology*, 54(3), 235–259.

Berchtold, A. (2003). Mixture Transition Distribution (MTD) Modeling of Heteroscedastic Time Series. *Computational Statistics and Data Analysis*, 41(3-4), 399–411.

Berglund, E. & Brannas, K. (2001). Plants' entry and exit in Swedish municipalities. *Annals of Regional Science*, 35(3), 431–448.

Berry, P., Dawson, T., Harrison, P., & Pearson, R. (2002). Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Global Ecology and Biogeography*, 11(6), 453–462.

Biggeri, A., Dreassi, E., Lagazio, C., & Böhning, D. (2003). A Transitional Non-parametric Maximum Pseudo-likelihood Estimator for Disease Mapping. *Computational Statistics and Data Analysis*, 41(3-4), 617–629.

Black, L., David, R., Brouillete, R., & Hunt, C. (1986). Effects of birth weight and ethnicity on incidence of SIDS. *Journal of Pediatrics*, 108, 209–214.

Blair, P., Fleming, P., Smith, I., Platt, M., Young, J., Nadin, P., Berry, P., Golding, J., & the CESDI SUDI research group (1999). Babies sleeping with parents: case-control study of factors influencing the risk of the sudden infant death syndrome. *British Medical Journal*, 319, 1457–1462.

Blair, P., Nadin, P., Cole, T., Fleming, P., Smith, I., Platt, M., Berry, P., & Golding, J. (2000). Weight gain and sudden infant death syndrome: changes in weight z-scores may identify infants at increased risk. *Archives of Disease in Childhood*, 82, 462–469.

Blundell, R., Griffith, R., & Windmeijer, F. (1995). Individual Effects and Dynamics in Count Data Models. Working Paper W95/15. Institute for Fiscal Studies, London.

Bockenholt, U. (1999). Analysing Multiple Emotions Over Time by Autoregressive Negative Multinomial Regression Models. *J. Am. Statist. Ass.*, 94, 757–765.

Bockenholt, U. (2003). Analysing State Dependences in Emotional Experiences by Dynamic Count Data Models. *Appl. Statist.*, 52(2), 213–226.

Böhning, D. (1999). *Computer-Assisted Analysis of Mixtures and Applications: Meta-analysis, disease mapping and others*. Number 81 in Monographs on Statistics and Applied Probability. London: Chapman and Hall.

Böhning, D. & Ayuthya, R. (1999). Analysis of geographical heterogeneity in live-birth ratios in Thailand. *Journal of Epidemiology and Biostatistics*, 4(2), 115–122.

Böhning, D., Dietz, E., & Schlattmann, P. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society A*, 162(2), 195–209.

Böhning, D., Dietz, E., & Schlattmann, P. (2000). Space-time mixture modelling of public health data. *Statistics in Medicine*, 19, 2333–2344.

Böhning, D. & Seidel, W. (2003). Editoral: Recent Developments in Mixture Models. *Computational Statistics and Data Analysis*, 41(3-4), 349–357.

Bollerslev, T., Chou, R., & Kroner, K. (1992). ARCH Modeling in Finance. A Review of the Theory and Empirical Evidence. *J. Econom.*, 52, 5–59.

Booth, J., Casella, G., Friedl, H., & Hobert, J. (2003). Negative binomial loglinear mixed models. *Statistical Modelling*, 3, 179–191.

Borman, B., Fraser, J., & de Boer, G. (1988). A national study of sudden infant death syndrome in New Zealand. *New Zealand Medical Journal*, 101(848), 413–415.

Boswell, M. & Patil, G. (1970). Chance mechanisms generating the negative binomial distributions. In G. Patil (Ed.), *Random Counts in Models and Structures*, volume 1 (pp. 3–22). University Park, PA: Pennsylvania State University Press.

Braga, A., Zanobetti, A., & Schwartz, J. (2002). The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities. *Environmental Health Perspectives*, 110(9), 859–863.

Brannas, K. (1994). Estimating and Testing in Interger-valued AR(1) Models. Working Paper No. 335, Department of Economics, University of Umea, Sweden.

Brockwell, P. & Davis, R. (2002). *Introduction to Time Series and Forecasting*. New York: Springer-Verlag, second edition.

Brown, H. & Prescott, R. (1999). *Applied Mixed Models in Medicine*. Statistics in Practice. New York: John Wiley & Sons, Inc.

Brumback, B., Ryan, L., Schwartz, J., Neas, L., Start, P., & Burge, H. (2000). Transitional regression models, with application to environmental time series. *Journal of the American Statistical Association*, 95(449), 16–27.

Burr, I. (1974). *Applied Statistical Methods*. New York: Academic Press.

Byard, R., Becker, L., Berry, P., Campbell, P., Fitzgerald, K., Hilton, J., Krous, H., & Rognum, T. (1996). The pathological approach to sudden infant death - consensus or confusion? Recommendations from the second SIDS Global Strategy Meeting, Stravanger, Norway, and the Third Australasian SIDS Global Strategy Meeting, Gold Coast, Australia. *American Journal of Forensic Medicine*, 17(2), 103–105.

Cameron, A. & Trivedi, P. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Campbell, M. (1989). Sudden infant death syndrome and environmental temperature: further evidence for a time-lagged relationship. *The Medical Journal of Australia*, 151, 365–367.

Campbell, M. (1994). Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *Journal of the Royal Statistical Society A*, 157(2), 191–208.

Campbell, M., Julious, S., Peterson, C., & Tobias, A. (2001). Atmospheric pressure and sudden infant death syndrome in Cook County, Chicago. *Paediatric and Perinatal Epidemiology*, 15, 287–289.

Campbell, M. & Tobias, A. (2000). Causality and temporality in the study of short-term effects of air pollution on health. *International Journal of Epidemiology*, 29, 271–273.

Carinanos, P., Galan, C., Alcazar, P., & Dominguez, E. (2000). Meteorological phenomena affecting the presence of solid particles suspended in the air during winter. *International Journal of Biometeorology*, 44(1), 6–10.

Chatfield, C. (1996). *The Analysis of Time Series, an Introduction.* Texts in Statistical Science. London: Chapman and Hall, fifth edition.

Cheung, Y. (2002). Zero-inflated models for regression analysis of count data: a study of growth and development. *Statistics in Medicine*, 21(10), 1461–1469.

Chilcoat, H. & Schuetz, C. (1996). Age-specific patterns of hallucinogen use in the US population: An analysis using generalized additive models. *Drug & Alcohol Dependence*, 43(3), 143–154.

Cliff, N. (1987). *Analyzing Multivariate Data.* USA: Harcourt Brace Jovanovich.

Coe, J. & Peterson, R. (1963). Sudden unexpected death in infancy and milk sensitivity. *Journal of Laboratory and Clinical Medicine*, 62, 477–480.

Cohen, A. (1963). Estimation in mixtures of discrete distributions. In *Proceedings of the International Symposium on Discrete Distributions* (pp. 373–378). Montreal.

Cooper, R. & Weekes, A. (1983). *Data, Models and Statistical Analysis.* Oxford: Philip Allan Publishers Ltd.

Cox, D. & Miller, H. (1977). *The Theory of Stochastic Processes.* London: Chapman and Hall, Science Paperback edition.

Dalrymple, M., Hudson, I., & Barnett, A. (2001). Survival, block bootstrap and mixture methods for detecting change points in discrete time series data with application to SIDS. In B. Klein & L. Karsholm (Eds.), *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling* (pp. 135–146).

Dalrymple, M., Hudson, I., & Ford, R. (2003). Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Computational Statistics and Data Analysis*, 41(3-4), 491–504.

Davis, R., Dunsmuir, W., & Wang, Y. (1999). *Asymptotics, Nonparametrics, and Time Series*, chapter Modeling Time Series of Count Data, (pp. 63–113). Marcel Dekker.

Davis, R., Dunsmuir, W., & Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, 87(3), 491–505.

Davison, A. & Hinkley, D. (1997). *Bootstrap Methods and Their Application.* Cambridge: Cambridge University Press.

Deacon, E., O'Reilly, M., & Williams, A. (1979). Some statistical and climatological aspects of the incidence of the sudden infant death syndrome. *Australian Paediatrics Journal*, 15, 248–254.

Deacon, E. & Williams, A. (1982). The incidence of the sudden infant death syndrome in relation to climate. *International Journal of Biometeorology*, 26(3), 207–218.

Dean, C. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association*, 87(418), 451–457.

Dean, C. & Lawless, J. (1989). Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association*, 84, 467–472.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum Likelihood Estimation from Incomplete data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177–188.

Dettling, M. & Buhlmann, P. (2003). Volatility and Risk Estimation with Linear and Non-Linear Methods Based on High Frequency Data. To appear in Applied Finacial Economics.

Diggle, P., Heagerty, P., Liang, K., & Zeger, S. (2002). *Analysis of Longitudinal Data.* Oxford Statistical Science Series. Oxford: Oxford University Press.

Dobbie, M. & Welsh, A. (2001). Modelling Correlated Zero-Inflated Count Data. *Australian & New Zealand Journal of Statistics*, 43(4), 431–444.

Dobson, A. (2002). *An Introduction to Generalized Linear Models.* London: Chapman and Hall, 2nd edition.

Dominici, F., McDermott, A., Zeger, S., & Samet, J. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology*, 156(3), 193–203.

Donoghue, A. & Thomas, M. (1999). Point source sulphur dioxide peaks and hospital presentations for asthma. *Occupational and Environmental Medicine*, 56(4), 232–236.

Douglas, A., Allan, T., & Helms, P. (1996). Seasonality and the sudden infant death syndrome during 1987—9 and 1991—3 in Australia and Britain. *British Medical Journal*, 312, 1381–1383.

Douglas, A., Helms, P., & Jolliffe, I. (1998). Seasonality of sudden infant death syndrome in mainland Britain and Ireland 1985—95. *Archives of Disease in Childhood*, 79, 269–270.

Dunlop, S. (2001). *A Dictionary of Weather.* Oxford University Press.

Dunson, D. & Haseman, J. (1999). Modeling tumour onset and multiplicity using transition models with latent variables. *Biometrics*, 55, 965–970.

Durlach, J., Pages, N., Bac, P., Bara, M., & Guiet-Bara, A. (2002). Magnesium deficit and sudden infant death syndrome (SIDS): SIDS due to magnesium deficiency and SIDS due to various forms of magnesium depletion: possible importance of the chronopathological form. *Magnesium Research*, 15(3-4), 269–278.

Dwyer, T. & Ponsonby, A. (1992). Sudden infant death syndrome — insights from epidemiological research. *Journal of Epidemiology and Community Health*, 46, 98–102.

Dwyer, T. & Ponsonby, A. (1995). SIDS epidemiology and incidence. *Pediatric Annals*, 24, 350–356.

Egan, V. & Angus, S. (2004). Is social dominance a sex-specific strategy for infidelity? *Personality and Individual Differences*, 36, 575–586.

Engelberts, A. (1991). *Cot Death in the Netherlands: An epidemiological study.* Amsterdam, The Netherlands: VU University Press.

Environment Canada (2002). Wind Chill Program. *http://www.msc.ec.gc.ca/education/windchill/index_e.cfm.*

Erbas, B. & Hyndman, R. (2001). Statistical methodological issues in studies of air pollution and respiratory disease. In B. Klein & L. Karsholm (Eds.), *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling* (pp. 179–186).

Everitt, B. S. & Hand, D. (1981). *Finite Mixture Distributions.* London, New York: Chapman and Hall.

Faddy, M. (1996). Stochastic models for analysis of species abundance data. *Statistics in Ecology and Environmental Monitoring*, (pp. 33–40).

Faddy, M. (1997). Extended Poisson process modelling and analysis of count data. *Biometrical Journal*, 39(4), 431–440.

Feuerverger, A. (1979). On some methods of analysis for weather experiments. *Biometrika*, 66(3), 655–658.

Fisher, G. & Taylor, K. (1999). Cold comfort for coal in Christchurch. *Aniwaniwa*, 11.

Fitzgerald, K. (2000). International SIDS numbers. (compiled by), Chairman, SIDS Global Strategy Task Force, kaarene@sidsaustralia.org.au.

Fleming, P., Blair, P., Bacon, C., Bensley, D., Smith, I., Taylor, E., Berry, P., Golding, J., & Tripp, J. (1996a). Environment of infants during sleep and risk of sudden infant death syndrome: results of 1993—1995 case-control study for confidential inquiry into still births and deaths in infancy. *British Medical Journal*, 313, 191–195.

Fleming, P., Blair, P., Pollard, K., Platt, M., Leach, C., Smith, I., Berry, P., & Golding, J. (1996b). Pacifier use and the sudden infant death syndrome: results from the CESDI/SUDI case-control study. *British Medical Journal*, 313, 191–195.

Ford, R. (1986). Postneonatal mortality in Christchurch. *New Zealand Medical Journal*, 99(815), 939–941.

Ford, R., McCormick, H., Pearce, G., & Harnet, P. (1990). Cot deaths in Canterbury: the pattern over twenty years. *New Zealand Medical Journal*, 103, 588.

Ford, R., Schluter, P., Mitchell, E., Taylor, B., Scragg, R., Stewart, A., & the New Zealand Cot Death Study Group (1998). Heavy caffeine intake in pregnancy and sudden infant death syndrome. *Archives of Disease in Childhood*, 78, 9–13.

Ford, R., Schluter, P., Taylor, B., Mitchell, E., Scragg, R., & the members of the New Zealand Cot Death Study Group (1996). Allergy and the risk of sudden infant death syndrome. *Clinical and Experimental Allergy*, 26, 580–584.

Ford, R., Taylor, B., Mitchell, E., Enright, S., Stewart, A., Becroft, D., R.Scragg, Hassall, I., Barry, D., Allen, E., & Roberts, A. (1993). Breastfeeding and the risk of sudden infant death syndrome. *International Journal of Epidemiology*, 22(5), 885–890.

Froggatt, P., Lynas, M., & Marshall, T. (1971). Sudden unexpected death in infants ("cot death"). *Ulster Medical Journal*, 40, 116–135.

Fryer, R. & Nicholson, M. (2002). Assessing covariate-dependent contaminant time-series in the marine environment. *ICSE Journal of Marine Science*, 59(1), 1–14.

Fujisaki, J., van Belle, G., & Sampson, P. (1990). 'Size' and 'shape' variables in the presence of covariates: an application to the SIDS. *Journal of Clinical Epidemiology*, 43, 173–180.

George, M., Wiklund, L., Aastrup, M., Pousette, J., Thunholm, B., Saldeen, T., Wernroth, L., Zaren, B., & Holmberg, L. (2001). Incidence and geographical distribution of sudden infant death syndrome in relation to content of nitrate in drinking water and groundwater levels. *European Journal of Clinical Investigation*, 31, 1083–1094.

Gilman, E., Cheng, K., Winter, H., & Scragg, R. (1995). Trends in rates and seasonal distribution of sudden infant deaths in England and Wales, 1988-92. *British Medical Journal*, 310, 631–2.

Glasbey, C. (2001). Nonlinear Autoregressive Time Series with Multivariate Gaussian Mixtures as Marginal Distributions. *Applied Statistics*, 50, 143–154.

Gold, E. & Godek, G. (1961). Antibody to milk in serum of normal infants and infants who died suddenly and unexpectedly (abstract). *American Journal of Diseases in Childhood*, 102, 542.

Goldberg, J., Hornung, R., Uamashita, T., & Wehrmacher, W. (1986). Age at death and risk factors in SIDS. *Australian Paediatrics Journal*, 22(Supplement), 21–28.

Goldberg, M., Burnett, R., Valois, M., Flegel, K., Bailar, J., Brook, J., Vincent, R., & Radon, K. (2003). Associations between ambient air pollution and daily mortality among persons with congestive heart failure. *Environmental Research*, 91(1), 8–20.

Golding, J., Limerick, S., & Macfarlane, A. (1985). *Sudden Infant Death. Patterns, Puzzles and Problems.* Somerset, England: Open Books Publishing Ltd.

Golub, G. & Loan, C. V. (1996). *Matrix Computation.* Johns Hopkins studies in the mathematical sciences. Baltimore: Johns Hopkins University Press, 3rd edition.

Gordon, A., Ahmer, O. E., Chan, R., Madani, O. A., Braun, J., Weir, D., Busuttil, A., & Blackwell, C. (2002a). Why is smoking a risk factor for sudden infant death syndrome? *Child: Care, Health and Development*, 28(Supplement 1), 23–25.

Gordon, A., MacKenzie, D., Ahmer, O. E., Madani, O. A., Braun, J., Weir, D., Busuttil, A., & Blackwell, C. (2002b). Evidence for a genetic component in sudden infant death syndrome. *Child: Care, Health and Development*, 28(Supplement 1), 27–29.

Gozal, D., Lorey, F., Chandler, D., Derry, M., Lisbin, A., Keens, T., Cunningham, G., & Ward, S. (1994). Incidence of sudden infant death syndrome in infants with sickle cell trait. *Journal of Pediatrics*, 124, 211–214.

Grech, V., Savona-Ventura, C., Agius-Muscat, H., & Janulova, L. (2003). Seasonality of births is associated with seasonality of marriages in Malta. *Journal of Biosocial Science*, 35(1), 95–105.

Greene, W. (2000). *Econometric Analysis.* Upper Saddle River, New Jersey: Prentice Hall, 4th edition.

Grether, J. & Schulman, J. (1989). SIDS and birth weight. *Journal of Pediatrics*, 114, 561–567.

Grim, J. & Haindl, M. (2003). Texture Modelling by Discrete Distribution Mixtures. *Computational Statistics and Data Analysis*, 41(3-4), 603–615.

Grunwald, G., Hyndman, R., Tedesco, L., & Tweedie, R. (2000). Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics*, (pp. 479–495).

Grunwald, G. & Jones, R. (2000). Markov models for time series with mixed distribution. *Environmetrics*, 11.

Guntheroth, W. (1995). *Crib death: the sudden infant death syndrome.* Armonk, New York: Futura Publishing Co., Inc., 3 edition.

Guthrie, D., Birnbaum, Z., Dixon, W., Fienberg, S., Gentleman, J., Landwehr, J., Mann, N., Mielke, P., Neter, J., Roberts, D., Ryzin, J. V., Tamura, H., Wilburn, A., & of Nonstandard Mixtures of Distributions, J. W. P. (1989). Statistical Models and Analysis in Auditing. *Statist. Sci.*, 4, 2–33.

Ha, E., Cho, S., Park, H., Chen, D., Chen, C., Wang, L., Xu, X., & Christiani, D. (2002). Does standing at work during pregnancy result in reduced infant birth weight? *Journal of Occupation and Environmental Medicine*, 44(9), 815–821.

Haglund, B. & Cnattingius, S. (1990). Cigarette smoking as a risk factor for sudden infant death syndrome: a population-based study. *American Journal of Public Health*, 80, 29–32.

Hajat, S. & Haines, A. (2002). Associations of cold temperatures with GP consultations for respiratory and cardiovascular disease amongst the elderly in London. *International Journal of Epidemiology*, 31(4), 825–830.

Hall, D. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56, 1030–1039.

Härdle, W. (1990). *Applied Nonparametric Regression.* Econometric Society Monographs. Cambridge: Cambridge University Press.

Harman, H. (1976). *Modern Factor Analysis.* Chicago: The University of Chicago Press, 3rd edition.

Harper, R. & Hoffman, J., Eds. (1988). *Sudden Infant Death Syndrome. Risk Factors and Basic Mechanisms.* PMA Publishing Corporation: New York.

Harrell, F. (2001). *Regression Modeling Strategies With Application to Linear Models, Logistic Regression, and Survival Analysis.* Springer Series in Statistics. New York: Springer.

Hastie, T. & Tibshirani, R. (1990). *Generalized Additive Models.* Monographs on Statistics and Applied Probability; 43. London: Chapman and Hall.

Hauck, F., Moore, C., Herman, S., Donovan, M., Kalelkar, M., Christoffel, K., Hoffman, H., & Rowley, D. (2002). The contribution of prone sleeping position to the racial disparity in sudden infant death syndrome: the Chicago infant mortality study. *Pediatrics*, 110(4), 772–780.

Heinrich, J., Bolte, G., Holscher, B., Douwes, J., Lehmann, I., Fahlbusch, B., Bischof, W., Weiss, M., Borte, M., Wichmann, H., & LISA Study Group (2002). Allergens and endotoxin on mothers' mattresses and total immunoglobulin E in cord blood of neonates. *European Respiratory Journal*, 20(3), 617–623.

Hoffmann, H., Hunter, J., Ellish, N., Janerich, D., & Goldberg, J. (1988). *SIDS Risk Factors and Basic Mechanisms*, chapter Adverse reproductive factors and SIDS, (pp. 153–172). PMA Publishing Corp.: New York.

Hosmer, D. & Lemeshow, S. (1980). A goodness-of-fit test for the multiple logistic regression model. *Communications in Statistics*, A10, 1043–1069.

Hosmer, D. & Lemeshow, S. (2000). *Applied Logistic Regression.* Wiley Series in Probability and Statistics. New York: Wiley, 2nd edition.

Huerta, G., Jiang, W., & Tanner, M. (2003). Time Series Modeling via Hierarchical Mixtures. *Statistica Sinica*, 13, 1097–1118.

Hyndman, R. & Grunwald, G. (2000). Generalized additive modelling of mixed distribution Markov models with application to Melbourne's rainfall. *Australian and New Zealand Journal of Statistics*, 42(2), 145–158.

Ibrahim, J., chen, M., & Ryan, L. (2000). Bayesian variable selection for time series count data. *Statistica Sinica*, 10(3), 971–987.

Iyasu, S., Randall, L., Welty, T., Hsia, J., Kinney, H., Mandell, F., McClain, M., Randall, B., Habbe, D., Wilson, H., & Willinger, M. (2002). Risk factors for sudden infant death syndrome among northern plains Indians. *Journal of the American Medical Association*, 288(21), 2717–2723.

Jewkes, R., Levin, J., & Penn-Kekana, L. (2002). Risk factors for domestic violence: findings from a South African cross-sectional study. *Social Science in Medicine*, 55(9), 1603.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* London: Chapman and Hall.

Johnson, N. & Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions.* Boston: Houghton Mifflin.

Jolliffe, I. (2002). *Principal Component Analysis.* Springer Series in Statistics. New York: Springer, second edition.

Jones, M., Ponsonby, A., Dwyer, T., & Gilbert, N. (1994). The relation between climatic temperature and sudden infant death syndrome differs among communities: results from an ecologic analysis. *Epidemiology*, 5, 332–336.

Jones, P. & McLachlan, G. (1991). Fitting mixture distributions to phenylthiocarbamide (PTC) sensitivity. *American Journal of Human Genetics*, 48, 117–120.

Jordan, M. & Jacobs, R. (1994). Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Comp.*, 6, 181–214.

Julious, S. (1997). There is still seasonality in sudden infant death syndrome in England and Wales. *Journal of Epidemiology and Community Health*, 51, 101–102.

Kassomenos, P., Gryparis, A., Samoli, E., Katsouyanni, K., Lykoudis, S., & Flocas, H. (2001). Atmospheric circulation types and daily mortality in Athens, Greece. *Environmental Health Perspectives*, 109(6), 591–596.

Katz, R. (1977). Precipitation as a Chain Dependent Process. *J. Appl. Meteor.*, 16, 671–676.

Kedem, B. & Fokianos, K. (2002). *Regression Models for Time Series Analysis.* Wiley Series in Probability and Statistics. New Jersey: John Wiley & Sons Inc.

Kim, D. & Kon, S. (1994). Alternative Models for the Conditional Heteroscedasticity of Stock Returns. *J. Business*, 67, 563–598.

Kim, J. & Mueller, C. (1978). *Factor Analysis, Statistical Methods and Practical Issues.* Quantitative Applications in the Social Sciences. London: Sage Publications.

Kivimaki, M., Elovainio, M., Vahtera, J., & Ferrie, J. (2003). Organisational justice and health of employees: prospective cohort study. *Occupational and Environmental Medicine*, 69(1), 27–33.

Kleemann, W., Schlaud, M., Poets, D., Rothämel, T., & Tröger, J. (1996). Hyperthermia in sudden infant death. *International Journal of Legal Medicine*, 109, 139–142.

Kleinbaum, D. (1996). *Survival Analysis: a self-learning text.* New York: Springer.

Knight, B. (1983). *Sudden Death in Infancy — the Cot Death Syndrome.* London: Faber and Faber.

Knöbel, H., Chen, D., & Liang, K. (1995). Sudden infant death syndrome in relation to weather and optimetrically measured air pollution in Taiwan. *Pediatrics*, 96(6), 1106–1110.

Knöbel, H., Yang, W., & Ho, M. (1994). Urban-rural and regional differences in infant mortality in Taiwan. *Social Science in Medicine*, 39(6), 815–822.

Knox, E. & Lancashire, R. (1991). Cot deaths in Birmingham. *Journal of Public Health in Medicine*, 13, 142–150.

Kokudo, N., Vera, D., Tada, K., Koizumi, M., Seki, M., Matsubara, T., Ohta, H., Yamaguchi, T., Takahashi, T., Hakajima, T., & Muto, T. (2002). Predictors of successful hepatic resection: prognostic usefulness of hepatic asialoglycoprotein receptor analysis. *World Journal of Surgery*, 26(11).

Kraus, J., Greenland, S., & Bulterys, M. (1989). Risk factors for sudden infant death syndrome in the US collaborative perinatal project. *International Journal of Epidemiology*, 18, 113–120.

Lachenburch, P. (1976). Analysis of Data with Clumping at Zero. *Biometrische Zeitschrift*, 18, 351–356.

Lachenburch, P. (1992). Utility of Regression Analysis in Epidemiologic studies of the Elderly. In R. Wallace & R. Woolson (Eds.), *The Epidemologic Study of the Elderly.* Oxford University Press.

Lam, D. & Miron, J. (1994). Global patterns of seasonal variation in human fertility. *Annals of the New York Academy of Sciences*, 709, 9–28.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1–14.

Le, N., Martin, R., & Raftery, A. (1996). Modelling Flat Stretches, Bursts, and Outliers in time series using Mixture Transition Distribution Models. *J. Amer. Statist. Assoc.*, 91, 1504–1515.

Leach, C., Blair, P., Fleming, P., Smith, I., Platt, M., Berry, P., Golding, J., & the CESDI SUDI Research Group (1999). Epidemiology of SIDS and Explained Sudden Infant Deaths. *Pediatrics*, 104(4).

Lee, A., Stevenson, M., Wang, K., & Yau, K. (2002). Modeling young driver motor vehicle crashes: data with extra zeros. *Accident; Analysis and Prevention*, 34(4), 515–521.

Lee, N., Chan, Y., Davies, D., Lau, E., & Yip, D. (1989). Sudden infant death syndrome in Hong Kong: confirmation of low incidence. *British Medical Journal*, 298(6675), 721.

Lemeshow, S. & Hosmer, D. (1982). The use of goodness-of-fit statistics in the development of logistic regression models. *American Journal of Epidemiology*, 115, 92–106.

L'Hoir, M., Engelberts, A., van Well, G., Westers, P., Mollenbergh, G., Wolters, W., & Huber, J. (1998). Case-control study of current validity of previously described risk factors for SIDS in the Netherlands. *Archives of Disease in Childhood*, 79, 386–393.

Linacre, E. & Geerts, B. (1997). *Climates and Weather Explained.* London: Routledge.

Lindenmayer, D., Cunningham, R., MacGregor, C., Incoll, R., & Michael, D. (2003). A survey design for monitoring the abundance of arboreal marsupials in the Central Highlands of Victoria. *Biological Conservation*, 110(1), 161–167.

Lindsay, B. (1995). Mixture Models: Theory, Geometry, and Applications. NSF-CBMS Regional Conference in Probability and Statistics, Vol.5 Institute of Mechanical Statistics, Hayward.

Lipfert, F., Zhang, J., & Wyzga, R. (2000). Infant mortality and air pollution: a comprehensive analysis of U.S. data for 1990. *Journal of the Air and Waste Management Association*, 50, 1350–1366.

Lyon, T., Partriarca, M., Howatson, G., Fleming, P., Blair, P., & Fell, G. (2002). Age dependence of potentially toxic elements (Sb, Cd, Pb, Ag) in human liver tissue from paediatric subjects. *Journal of Environmental Monitoring*, 4(6), 1034–1039.

MacDonald, I. & Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series.* London: Chapman and Hall.

Macey, P., Schluter, P., & Ford, R. (2000). Weather and the risk of sudden infant death syndrome: the effect of wind. *Journal of Epidemiology and Community Health*, 54, 333–339.

Manly, B. (1986). *Multivariate Statistical Methods, A Primer.* London: Chapman and Hall.

Manning, W., Duan, N., & Rogers, W. (1987). Monte Carlo Evidence on the Choice Between Sample Selection and Two-part Models. *Journal of Econometrics*, 35, 59–82.

Mardia, K. (1972). *Statistics of Directional Data.* London: Academic Press.

Matthews, T. (2002). Sudden infant death syndrome — a defect in circulatory control? *Child: Care, Health and Development*, 28(Supplement 1), 41–43.

Matthews, T. & O'Brien, S. (1985). Perinatal epidemiological characteristics of the sudden infant death syndrome in an Irish population. *Irish Medical Journal*, 78, 251–253.

McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models.* London: Chapman and Hall, 2nd edition.

McCulloch, C. & Searle, S. (2001). *Generalized, Linear, and Mixed Models.* Wiley Series in Probability and Statistics. New York: John Wiley & Sons, Inc.

McGlashan, N. & Grice, A. (1983). Sudden infant deaths and seasonality in Tasmania, 1970—1976. *Social Science in Medicine*, 17(13), 885–888.

McIlveen, J. (1986). *Basic Meteorology: a Physical Outline.* Berkshire, England: Van Nostrand Reinhold (UK) Co. Ltd.

McKenzie, D. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, 20, 822–835.

McLachlan, G. & Basford, K. (1988). *Mixture Models. Inference and Applications to Clustering.* New York: Marcel Dekker.

McLachlan, G., Bean, R., & Peel, D. (2001). EMMIX-GENE: A Mixture Model-based Program for the Clustering of Microarray Expression Data. Technical Report.

McLachlan, G. & Peel, D. (2000). *Finite Mixture Models.* Wiley Series in Probability and Statistics, Applied Probability and Statistics Section. John Wiley & Sons, Inc.

McLachlan, G., Peel, D., & Bean, R. (2003). Modelling High-Dimensional Data by Mixture of Factor Analysers. *Computational Statistics and Data Analysis*, 41(3-4), 379–388.

Milerad, J., Norvenius, G., & Wennergren, F. (1993). SIDS outdoors and seasonality in Sweden 1975—1987. *Acta Paediatrica*, 82, 1039–1042.

Miric, D. & Rumboldt, Z. (1993). The impact of meteorological factors on the onset of myocardial infarction in the coastal region of middle Dalmatia. *Giornale Italiano di Cardiologia*, 23, 655–660.

Mitchell, E. (1993). Sleeping position of infants and the sudden infant death syndrome. *Acta Paediatrica*, 389(Supplement), 387–389.

Mitchell, E., Brunt, J., & Everard, C. (1994). Reduction in mortality from sudden infant death syndrome in New Zealand: 1986-92. *Archives of Disease in Childhood*, 70, 291–294.

Mitchell, E., Clements, M., Williams, S., Stewart, A., Cheng, A., Ford, R., & the New Zealand Cot death study group (1999). Seasonal differences in risk factors for sudden infant death syndrome. *Acta Paediatrica*, 88, 253–258.

Mitchell, E., Ford, R., Taylor, B., Stewart, A., Becroft, D., Scragg, R., Barry, D., Allen, E., Roberts, A., & Hassall, I. (1992a). Further evidence supporting a causal relationship between prone sleeping position and SIDS. *Journal of Paediatric and Child Health*, 28(Supplement. 1), S9–S12.

Mitchell, E., Scragg, R., Stewart, A., Becroft, D., Hassall, I., Ford, R., Taylor, B., Roberts, A., Barry, D., & Allen, E. (1989). Cot death, the New Zealand study. *Contemporary Health Issues National Health Statistics Centre*, 1, 67–72.

Mitchell, E., Stewart, A., & Cowan, S. (1992b). Sudden infant death syndrome and weather temperature. *Paediatric and Perinatal Epidemiology*, 6, 19–28.

Mitchell, E., Stewart, A., Scragg, R., Ford, R., Taylor, B., Becroft, D., Thompson, J., Hassall, I., Barry, D., Allen, E., & Roberts, A. (1993a). Ethnic differences in mortality from sudden infant death syndrome in New Zealand. *British Medical Journal*, 306, 13–16.

Mitchell, E., Taylor, B., Ford, R., Stewart, A., Becroft, D., Thompson, J., Scragg, R., Hassall, I., Barry, D., Allen, E., & Roberts, A. (1992c). Four modifiable and other major risk factors for cot death: the New Zealand study. *Journal of Paediatric and Child Health*, 28(Supplement 1), S3–S8.

Mitchell, E., Taylor, B., Ford, R., Stewart, A., Becroft, D., Thompson, J., Scragg, R., Hassall, I., Barry, D., Allen, E., & Roberts, A. (1993b). Dummies and the sudden infant death syndrome. *Archives of Disease in Childhood*, 68, 501–504.

Mood, A., Graybill, F., & Boes, D. (1974). *Introduction to the Theory of Statistics.* Statistics Series. Singapore: McGraw-Hill International Editions.

Moon, R. & Omron, R. (2002). Determinants of infant sleep position in an urban population. *Clinical Pediatrics*, 41(8), 569–573.

Mooney, J., Helms, P., & Jolliffe, I. (2003). Fitting Mixtures of Von Mises Distributions: A Case Study Involving Sudden Infant Death Syndrome. *Computational Statistics and Data Analysis*, 41(3-4), 505–513.

Morgan, B., Finan, A., Yarnold, R., Petersen, S., Horsfield, M., Rickett, A., & Wailoo, M. (2002). Assessment of infant physiology and neuronal development using magnetic resonance imaging. *Child: Care, Health and Development*, 28(Supplement 1), 7–10.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33, 341–365.

Naeye, R. (1973). Pulmonar arterial abnormalities in the SIDS. *New England Journal of Medicine*, 289, 1167–1170.

National Health Statistics Centre (New Zealand) (1986). *Fetal and Infant Deaths.* New Zealand Health statistics report. Wellington.

National Health Statistics Centre (New Zealand) (1997). *Fetal and infant deaths.* New Zealand Health statistics report. Wellington.

National Health Statistics Centre (New Zealand) (1998). *Fetal and infant deaths.* New Zealand Health statistics report. Wellington: Wiley.

Nelson, E. (1996). *Sudden Infant Death Syndrome & Child Care Practices.* Hong Kong: Dr Tony Nelson, 2nd edition.

Nelson, E. & Taylor, B. (1988). Climatic and social associations with postneonatal mortality rates within New Zealand. *New Zealand Medical Journal*, 101, 443–446.

Nelson, E. & Taylor, B. (1999). International child care practices study: methods and study population. *Early Human Development*, 55(2), 149–168.

Nelson, E. & Taylor, B. (2001). International child care practices study: infant sleep poisition and parental smoking. *Early Human Development*, 64(1), 7–20.

Nelson, E., Taylor, B., Jenik, A., Vance, J., Walmsley, K., Pollard, K., Freemantle, M., Ewing, D., Einspieler, C., Engele, H., Ritter, P., Hides-Ripstein, G., Arancibia, M., Ji, X., Li, H., Bedard, C., Helweg-Larsen, K., Sidenius, K., Karlqvist, S., Poets, C., Barko, E., Kiberd, B., McDonnell, M., Donzelli, G., Piumelli, R., Landini, L., Giustardi, A., Nishida, H., Fukui, S., Sawaguchi, T., Ino, M., Horiuchi, T., Oguchi, K., Williams, S., Perk, Y., Tappin, D., Milerad, J., Wennborg, M., Aryayev, N., & Nepomyashchaya, V. (2001). International child care practices Study: infant sleeping environment. *Early Human Development*, 62(1), 43–55.

Nevid, J., Rathus, S., & Greene, B. (2000). *Abnormal Psychology in a Changing World.* Upper Saddle River, New Jersey: Prentice Hall, 4th edition.

New Zealand Meteorological Service (1961). Relative humidity table; prepared in the Meteorological Office, Wellington, New Zealand, 1944.

Newbould, M., Malam, J., McIllmuray, J., Morris, J., Telford, D., & Barson, A. (1989). Immunohistological localisation of staphylococcal toxic shock syndrome toxin (TSST-1) antigen in sudden infant death syndrome. *Journal of Clinical Pathology*, 42, 935–939.

Nieto, A., Ruiz-Ramos, M., Abdel-Kader, L., Conde, M., & Camacho, F. (2003). Gender differences in rising trends in cutaneous malignant melanoma in spain, 1975-98. *British Journal of Dermatology*, 148(1), 110–116.

Nishida, H. & Takashima, S. (2002). Quantification of trace elements in the brain of SIDS victims. *Forensic Science International*, 130(Supplement), S63–S64.

Office for National Statistics, UK (1998). Figures on SIDS and low birth weight released August 18 1998. *Lancet (quoted in)*, 352, 712.

Oh, M. & Lim, Y. (2001). Bayesian analysis of time series Poisson data. *Journal of Applied Statistics*, 28(2), 259–271.

Okanda, F., Dao, A., Njiru, B., Arija, J., Akelo, H., Toure, Y., Odulaja, A., Beier, J., Githure, J., Yan, G., Gouagna, L., Knols, B., & Killeen, G. (2002). Behavioural determinants of gene flow in malaria vector populations: Anopheles gambiae males select large females as mates. *Malaria Journal*, 1(1), 10.

Okcu, M., Goodman, K., Carozza, S., Weiss, N., Burau, K., Bleyer, W., & Cooper, S. (2002). Birth weight, ethnicity, and occurrence of cancer in children: a population-based, incident case-control study in the State of Texas, USA. *Cancer Causes Control*, 13(7), 595–602.

O'Neill, M. & Faddy, M. (2003). Use of binary and truncated negative binomial modelling in the analysis of recreational catch data. *Fisheries Research*, 60(2-3), 471–477.

Osius, G. & Rojek, D. (1992). Normal goodness-of-fit tests for multinomial models with large degrees-of-freedom. *Journal of the American Statistical Association*, 87, 1145–1152.

Oyen, M., Markstad, T., Skjaerven, R., Irgens, L., Helweg-Larsen, K., Alm, B., Norvenius, G., & Wennergren, G. (1997). Combined effects of sleeping position and prenatal risk factors in sudden infant death syndrome: the Nordic epidemiological SIDS study. *Pediatrics*, 100, 613–621.

Pace, R. (1998). Appraisal using generalized additive models. *Journal of Real Estate Research*, 15(1/2), 77–99.

Parish, W. & Barett, A. (1960). Hypersensitivity to milk and sudden death in infancy. *Lancet*, 2, 1106–1110.

Park, J. & Basawa, I. (2002). Estimation for mixtures of Markov processes. *Statistics & Probability Letters*, 59(3), 235–244.

Perneger, T. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal*, 316, 1236–1238.

Peterson, D. (1980). Evolution of the epidemiology of sudden infant death syndrome. *Epidemiologic Reviews*, 2, 97–112.

Peterson, D. (1989). *Sudden infant death syndrome. Medical aspects and psychological management*, chapter The epidemiology of sudden infant death syndrome, (pp. 3–17). Edward Arnold (Hodder and Stoughton): London.

Pinchbeck, G., P, D, C., Proudman, C., Morgan, K., Wood, J., & French, N. (2002). Risk factors and sources of variation in horse falls in steeplechase racing in the UK. *Preventive Veterinary Medicine*, 55(3), 179–192.

Piorecky, J., Becker, W., & Rose, M. (1997). Effect of Chinook winds on the probability of migraine headache occurrence. *Headache*, 37, 153–158.

Pollack, H. (2001). Sudden infant death syndrome, maternal smoking during pregnancy, and the cost-effectiveness of smoking cessation intervention. *American Journal of Public Health*, 91(3), 432–436.

Pollack, H. & Frohna, J. (2002). Infant sleep placement after the back to sleep campaign. *Pediatrics*, 109(4), 608–614.

Ponsonby, A., Jones, M., Lumley, J., Dwyer, T., & Gilbert, N. (1992). Sudden infant death syndrome: factors contributing to the difference in incidence between Victoria and Tasmania. *The Medical Journal of Australia*, 156, 252–254.

Price, C., Coope, I., & Byatt, D. (2002). A convergent variant of the Nelder-Mead algorithm. *Journal of Optimization Theory and Applications*, 113(1), 5–19.

Raftery, A. (1985). A Model for High-order Markov Chains. *J. Roy. Statist. Soc. B*, 47, 528–539.

Raftery, A. & Tavare, S. (1994). Estimation and Modelling Repeated Patterns in High Order Markov Chains with the Mixture Transition Distribution Model. *Appl. Statist.*, 43, 179–199.

Ramesh, N. & Davison, A. (2002). Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology*, 256(1-2), 106–119.

Raven, C. (1977). Crib deaths, sudden unexpected death in infancy, or the SIDS: a hypersensitivity reaction. *Journal of the American Medical Women's Association*, 32, 148–149.

Raven, C., Maverakis, N., Eveland, W., & Ackermann, W. (1978). The SIDS: a possible hypersensitivity reaction determined by distribution of IGG in lungs. *Journal of Forensic Science*, 23, 116–128.

Rawlings, J. (1988). *Applied Regression Analysis.* Probability Series. California: Wadsworth & Brooks/Cole Advanced Books & Software.

Reed, W., Larsen, C., Johnson, E., & MacDonald, G. (1998). Estimation of temporal variations in historical fire frequency from time-since-fire map data. *Forest Science*, 44(3), 465–475.

Reinsch, C. (1967). Smoothing by spline functions. *Numerische Mathematik*, 10, 177–183.

Reports on Public Health No. 113 (1965). Enquiry into sudden death in infancy.

Richardson, B. (1990). Cot mattress biodeteriation and SIDS. *Lancet*, 335, 670.

Rintahaka, P. & Hirvonen, J. (1986). The epidemiology of sudden infant death syndrome in Finland in 1969-1980. *Forensic Science International*, 30(2-3), 219–233.

Rothman, K. & Greenland, S. (1998). *Modern Epidemiology.* Philadelphia: Lippincott-Raven, 2nd edition.

Russell-Jones, D. (1985). Sudden infant death in history and literature. *Archives of Disease in Childhood*, 60, 278–281.

Sadeghi, H. (2003). Local or global asymmetry in gait of people without impairments. *Gait and Posture*, 17, 197–204.

Santee, W. (2002). Windchill index and military applications. *Aviation Space and Environmental Medicine*, 73(7), 699–702.

Schellscheidt, J., Oen, N., & Jorch, G. (1997). Interactions between maternal smoking and other prenatal risk factors for sudden infant death syndrome. *Acta Paediatrica*, 86, 857–863.

Schlattmann, P. (2003). Estimating the Number of Components in a Finite Mixture Model: The Special Case of Homogeneity. *Computational Statistics and Data Analysis*, 41(3-4), 441–451.

Schlattmann, P., Dietz, E., & B'ohning, D. (1996). Covariate adjusted mixture models and disease mapping with the program DismapWin. *Statistics in Medicine*, 15, 919–929.

Schluter, P., Ford, R., Brown, J., & Ryan, A. (1998). Weather temperatures and sudden infant death syndrome: a regional study over 22 years in New Zealand. *Journal of Epidemiology and Community Health*, 52, 27–33.

Schluter, P., Macey, P., & Ford, R. (2000). The relationship between inside and outside ambient temperatures in Christchurch, New Zealand. *Paediatric and Perinatal Epidemiology*, 14, 275–282.

Schwartz, J., Spix, C., Touloumi, G., Bachárová, L., Barumamdzadeh, T., le Tertre, L., Piekarksi, T., de Leon, A. P., Pönk—'a, A., Rossi, G., Saez, M., & Schouten, J. (1996). Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *Journal of Epidemiology and Community Health*, 50(Supplement 1), S3–S11.

Schwartz, P. (1997). Effects of metachlorophenylpiperazine infusions in patients with seasonal affective disorder and healthy control subjects. *Archives of General Psychiatry*, 54, 375–385.

Scragg, R., Mitchell, E., Stewart, A., Ford, R., Taylor, B. J., Hassall, I., Williams, S., & Thompson, J. (1996). Infant room-sharing and prone sleep position in sudden infant death syndrome. *Lancet*, 347, 7–12.

Social Report (1991). *Sudden infant death syndrome SIDS by states and territories, 1975 to 1988.* Tasmania: Australian Bureau of Statistics.

Spiers, P. (1990). Risk of SIDS by area of residence in the United States. *International Journal of Epidemiology*, 103, 45–48.

Stanton, A. (1984). Sudden infant death. Overheating and cot death. *Lancet*, 2, 1199–1201.

Steele, R. & Langworth, J. T. (1966). The relationship of antenatal and postnatal factors to sudden unexpected death in infancy. *Canadian Medical Association Journal*, 94, 1165–1171.

Steinschneider, A. (1972). Prolonged apnea and the SIDS: clinical and laboratory observations. *Pediatrics*, 50, 646–654.

Stern, R. & Coe, R. (1984). A Model Fitting Analysis of Daily Rainfall Data (with discussion). *J. Roy. Statist. Soc. Ser. A*, 147, 1–34.

Stieb, D., Judek, S., & Burnett, R. (2003). Meta-analysis of time-series studies of air pollution and mortality: update in relation to the use of generalized additive models. *Journal of Air & Waste Association*, 53(3), 258–261.

Stratoudakis, Y., Bernal, M., Borchers, D., & Borges, M. (2003). Changes in the distribution of sardine eggs and larvae off Portugal, 1985—2000. *Fisheries Oceanography*, 12(1), 49–60.

Succi, G., Pedrycz, W., Stefanovic, M., & Miller, J. (2003). Practical assessment of themodels for identification of defect-prone classes in object-oriented commercial systems using design metrics. *Journal of Systems and Software*, 65(1), 1–12.

Sullivan, F. & Barlow, S. (2001). Review of risk factors for sudden infant death syndrome. *Paediatric and Perinatal Epidemiology*, 15, 144–200.

Taube, A. (1986). Sensitivity, specificity and predictive values: a graphical approach. *Statistics in Medicine*, 5, 585–591.

Templeman, C. (1892). Two hundred and fifty-eight cases of suffocation of infants. *Edinburgh Medical Journal*, 38, 322–329.

Thanasoulias, N., Parisis, N., & Evmiridis, N. (12003). Multivariate chemometrics for the forensic discrimination of blue ball-point pen inks based on their Vis spectra. *Foresnic Science International*, 138, 75–84.

Thompson, H., Valdimarsdottir, H., Winkel, G., Jandorf, L., & Redd, W. (2004). The Group-Based Medical Mistrust Scale: psychometric properties and association with breast cancer screening. *Preventive Medicine*, 38, 209–218.

Thyregod, P., Carstensen, J., Madsen, H., & Arnbjerg-Nielsen, K. (1999). Integer valued autoregressive models for tipping bucket rainfall measurements. *Environmetrics*, 10(4), 395–411.

Titterington, D., Smith, A., & Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions.* New York: Wiley.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford: Clarendon Press.

Toomey, S. & Bernstein, H. (2001). Sudden infant death syndrome. *Current Opinion in Pediatrics*, 13, 207–210.

Tooze, J., Grunwald, G., & Jones, R. (1994). Analysis of Repeated Measures Data with Clumping at Zero. *Journal Occup. Health Safety*, 10, 205–211.

Tooze, J., Grunwald, G., & Jones, R. (2002). Analysis of Repeated Measures Data with Clumping at Zero. *Statistical Methods in Medical Research*, 11, 341–355.

Toscas, P. & Faddy, M. (2003). Likelihood based analysis of longitudinal count data using a generalised Poisson model. *Statistical Modelling*, 3, 99–108.

Toubas, P., Duke, J., McCaffree, M., Mattice, C., Bendell, D., & Orr, W. (1986). Effects of maternal smoking and caffeine habits on infantile apnea: a retrospective study. *Pediatrics*, 78, 159–163.

Trube-Becker, E. (1978). Enteral bacterial infection as a possible cause of cot death. *Forest Science*, 11, 171–174.

van Duijin, M. & Bockenholt, U. (1995). Mixture Models for the Analysis of Repeated Count Data. *Appl. Statist.*, 44, 473–485.

Verhoef, M., Rose, M., & Ramcharen, S. (1995). The relationship between Chinook conditions and women's physical and mental well-being. *International Journal of Biometeorology*, 38, 148–151.

Vix, J., Buguet, A., Straboni, S., & Beidari, H. (1987). [Sudden infant death and sickle cell anemia in the Sahel region of Africa] (French). *Medecine tropicale : revue du Corps de sante colonial*, 47, 153–159.

Wahba, G. & Wendelberger, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review*, 108, 1122–1145.

Wakley, T. (1855). Infants found dead in bed. *Lancet*, 1, 103.

Walsh, R., Tzelepis, F., Paul, C., & McKenzie, J. (2002). Environmental tobacco smoke in homes, motor vehicles and licensed premises: community attitudes and practices. *Australian and New Zealand Journal of Public Health*, 26(6), 536–542.

Wang, P. (2001). Markov zero-inflated Poisson regression models for a time series of counts with excess zeros. *Journal of Applied Statistics*, 28(5), 623–632.

Wang, P., Cockburn, I., & Puterman, M. (1998). Analysis of patent data — a mixed-Poisson-regression-model approach. *Journal of Business & Economic Statistics*, 16(1), 27–40.

Wang, P., Puterman, M., Cockburn, I., & Le, N. (1996). Mixed Poisson regression models with covariate dependent rates. *Biometrics*, 52, 381–400.

Wasley, A., Lepine, L., Jenkins, R., & Rubin, C. (2002). An investigation of unexplained infant deaths in houses contaminated with methyl parathion. *Environmental Health Perspectives*, 110(Supplement 6), 1053–1056.

Waymire, E. & Gupta, V. (1981). The Mathematical Structure of Rainfall Representation. 1. A Review of the Stochastic Rainfall Models. *Water Resources Res.*, 17, 1261–1272.

Weitoft, G., Hjern, A., Haglund, B., & Rosen, M. (2003). Mortality, severe morbidity, and injury in children living with single parents in Sweden: a population based study. *Lancet*, 361(9354), 289–295.

Wells, J. (1997). Can Risk Factors for Over-Heating Explain Epidemiological Features of Sudden Infant Death Syndrome? *Medical Hypotheses*, 48(2), 103–106.

Welsh, A., Cunningham, R., & Chambers, R. (2000). Methodology for Estimating the Abundance of Rare Animals: Seabird Nesting on North East Herald Cay. *Biometrics*, 56, 22–30.

Welsh, A., Cunningham, R., Donnelly, C., & Lindenmayer, D. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, 88, 297–308.

West, M. & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models.* New York: Springer-Verlag, 2 edition.

Wigfield, R., Gilbert, R., & Fleming, P. (1994). SIDS: risk reduction measures. *Early Human Development*, 38, 161–164.

Williams, A., Uren, E., & Bretherton, L. (1984). Respiratory viruses and sudden infant death. *British Medical Journal*, 288, 1491–1493.

Williams, S., Taylor, B., Ford, R., & Nelson, E. (1990). Growth velocity before sudden infant death. *Archives of Disease in Childhood*, 65, 1315–1318.

Winkelmann, R. (2000). *Econometric Analysis of Count Data.* New York: Springer, third edition.

Wisborg, K., Kewmodel, U., Henriksen, T., Olsen, S., & Secher, N. (2000). A prospective study of smoking during pregnancy and SIDS. *Archives of Disease in Childhood*, 83, 203–206.

Wolter, K. & Timlin, M. (1998). Measuring the strength of ENSO - how does 1997/98 rank? *Weather*, 53, 315–324.

Wong, C. & Li, W. (2000). On a Mixture Autoregressive Model. *J. Roy. Statist. Soc. B*, 62, 92–115.

World Health Organisation (1965). *Manual of the international statistical classification of diseases, injuries and causes of death: based on the recommendations of the Eighth Revision Conference, 1965, and adopted by the Nineteenth World Health Assembly*, volume 1. Geneva.

World Health Organisation (1977). *Manual of the international statistical classification of diseases, injuries and causes of death: based on the recommendations of the Ninth Revision Conference, 1975, and adopted by the Twenty-ninth World Health Assembly*, volume 1. Geneva.

World Health Organisation (2002a). Multicountry Evaluation: Integrated Management of childhood illness. *http://www.who.int/imci-mce/Methods/mortality.htm*.

World Health Organisation (2002b). World Health Organisation statistics — basic health indicators. *http://ww3.who.int/whosis/reported/reported.cfm?path=whosis,basic,reported*.

Yau, K., Lee, A., & Ng, A. (2002). A Zero-argumented Gamma Mixed Model for Longitudinal Data with Many Zeros. *Aust. N.Z. J. Stat.*, 44(2), 177–183.

Yau, K., Lee, A., & Ng, A. (2003). Finite Mixture Regression Model with Random Effects: Application to Neonatal Hospital Length of Stay. *Computational Statistics and Data Analysis*, 41(3-4), 359–366.

Yau, K. & McGilchrist, C. (1998). ML and REML Estimation in Survival Analysis with Time Dependent Correlated Frailty. *Statist. Med.*, 17, 1201–1213.

Zeger, S. (1988). A regression model for time series of counts. *Biometrika*, 75(4), 621–629.

Zeger, S. & Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, 44, 1019–1031.

Zink, P., Drescher, J., Verhagen, W., & Flik, J. (1987). Serological evidence of recent influenza virus A (H3N2) infections in forensic cases of the SIDS. *Archives of Virology*, 93, 223–232.

Zylke, J. (1989). Sudden infant death syndrome: resurgent research offers hope. *Journal of the American Medical Association*, 262, 1565–1566.